1969

Inference for components of variance models

Syed Taqi Mohammad Naqvi
Iowa State University

69-15,635

NAQVI, Syed Taqi Mohammad, 1921-
INFERENCE FOR COMPONENTS OF VARIANCE MODELS.

Iowa State University, Ph.D., 1969
Statistics

University Microfilms, Inc., Ann Arbor, Michigan

INFERENCE FOR COMPONENTS OF VARIANCE MODELS

by

Syed Taqi Mohammad Naqvi

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1969

TABLE OF CONTENTS

I. INTRODUCTION

Consider the model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$
$$(i = 1, \dots ,k \; ; \; j = 1, \dots ,n)$$

where

y_{ij} is the response of the jth unit in the ith group.

μ is the general mean.

$\mu + \alpha_i$ is the mean of units in the ith group. We presume that the terms α_i constitute a random sample from a normal population with zero mean and variance σ^2_α .

ϵ_{ij}'s are random variables presumed to have independent and identical normal distribution with zero mean and variance σ^2 . The distribution of ϵ_{ij} is independent of α_i for all i and j.

Such a model is called the one-way random effect model. The observation y_{ij} has variance $(\sigma^2+\sigma^2_\alpha)$ and σ^2 and σ^2_α are called the components of variance. Research workers wish to estimate variance components or some function of variance components, especially in genetic and sampling problems. For example, in a genetic experiment involving sires, presumed to be a random sample from some population of sires, one is interested in knowing the amount of sire-to-sire variation. In the study of wheat yield of West Pakistan, a pre-determined number of villages is selected at random from each administrative sub-unit called "taluka" and these fields are selected at random from each selected village. A sampler is interested in having an estimate of the ratio of village-to-village variation to the within village variation and of the intra-class correlation $\sigma^2_\alpha/(\sigma^2+\sigma^2_\alpha)$.

The examples are too numerous to give in the present study. We use the word "estimate" in a loose non-technical sense. In most applications of this type of model, the aim is to characterize in some way to what extent values for the parameters are consistent with the data and the model. This appears to be a general statement which encompasses the various inferential procedures that can be applied. There are several competitive procedures and the aim of this dissertation is to discuss these and to attempt a better understanding of them and to compare them on chosen sets of data.

We examine three approaches of the problem. (i) Bayesian estimation, (ii) likelihood inference, and (iii) goodness of fit inference. In Chapter II we have given a concise account of major contributions to the solution of problems from different points of view. In Chapter III we have considered the problem of estimating parametric functions of interest from the Bayesian point of view. We have used so-called non-informative priors which are independent of sample elements and have presented some solutions. We have used a reasonable informative prior based on the history of the case and the experience of the experimentor. The solution is presented as an illustration.

It is claimed by Barnard et al. (2) that the primary source of information should be the likelihood function. In Chapter IV, this aspect of the problem is investigated. The role played by the non-informative prior in a Bayesian inference is also investigated. We have used graphic method of illustration in this chapter.

In Chapter V we ask ourselves the question, given a set of data, of what values of the pair $(\sigma^2, \sigma_\alpha^2)$ are in consonant with the data at a desired level of significance, or what goodness of fit value can be attached to any particular pair $(\sigma^2, \sigma_\alpha^2)$. A concept of goodness of fit is defined and the problem is considered from this point of view. Numerical examples from the chosen set of data are presented and the goodness of fit results are compared with the likelihood results. Chapter VI contains a summary of results and conclusions.

II. REVIEW OF LITERATURE

The problem of drawing inference from the random effect model has been engaging the attention of statisticians and research workers for the past three decades or so. The problem has been looked upon from different points of view and solutions presented. Broadly, the work can be divided into the following main categories:

Non-Bayesian

a) Sampling Theory

b) Confidence Interval

c) Fiducial Limits

Bayesian

a) Posterior distribution of parameters of interest.

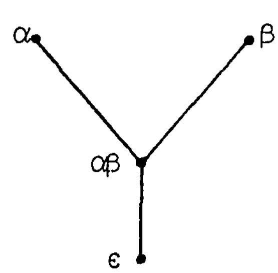
A brief review of work under each category with summary of main results is presented in this chapter.

A. Sampling Theory

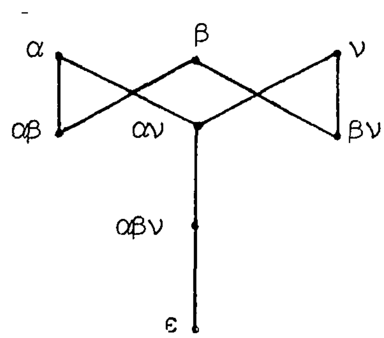
Earlier work on this approach to the problem has been in the direction of evolving suitable methods for the analysis of data in case of unequal sub-class numbers and the evaluation of the expectation of mean squares in each line of the analysis of variance table. Equal sub-class numbers present no difficulty. The procedure to estimate components of variance was to solve the simultaneous linear equations obtained by equating mean squares in the analysis of variance table with their respective expectations. Faced with a negative estimate, the solution suggested is to take the estimate of component concerned as zero. Some work has been done to estimate variance of estimates of components of variance in balanced designs. See Anderson (1), Crump (10,11), Daniel (12), Henderson (18), Kempthorne (22), Lucas (26), Satterthwaite (29), Snedecor (30), Wald (37,38), Wilk and Kempthorne (40), and Yates (42). Associated with this method are asymptotic formulae for estimating standard errors of estimators. The suggested procedure is to use estimate $\pm$ z . standard error, with z as a standard normal deviate, to indicate ranges of reasonable values for the parameters.

The problem of negative estimation was considered by Thompson, Jr. (32) and Thompson, Jr. and Moore (33). A component of variance with negative estimate is taken as zero and the estimates of other components are adjusted. They have developed an algorithm which they call the "pool-the-minimum-violator" algorithm. The method involves partial ordering of expected mean squares and a representation by a graph consisting of dots and lines. The method is applicable if the graph is a rooted tree. A rooted tree is defined as a graph with at least a point, called the root, such that from each point of the graph there is a unique path connecting the given point with the root. The utility is, therefore, restricted. For example in the following diagrams the method is applicable to (a) and (b), but not to (c).

(a)



(b)



(c)

The estimate, obtained by this process, for the one-way model is what Thompson, Jr (32) calls "restricted maximum likelihood estimate ".

On this problem of negative estimates, the empirical studies of Leone and Nelson (24) and Leone, Nelson and Johnson (25) are informative. A four-stage nested design was considered by them. In the balanced case, one thousand samples of size $5 \times 2 \times 2 \times 2 = 40$ each under eight different sets of population variances were analyzed. For each set, the error component is taken as unity and other components expressed as a ratio to the error component. Percentages of negative estimates of variance components were calculated. For normal populations, the results are reproduced in Table 1.

Similar results were obtained for unbalanced design, but here the percentage of negative estimates at one stage can be reduced at the cost of other stages through a suitable choice of sample size.

Obviously, these results are not directly applicable to one-way random model, but they give an indication. There is a reasonably good probability of getting a negative estimate of σ_α^2 from a single experiment even when in the population σ_α^2 is equal to or greater than σ^2 and moderate probability for higher values of σ_α^2 .

Klotz and Milton (23) have studied the mean square error of the following estimators of σ_α^2 for the balanced onefold classification.

1. Maximum likelihood estimator:

$$\hat{\sigma}_\alpha^2 = \frac{1}{n}\left[\frac{S_2}{k} - \frac{S_1}{k(n-1)}\right]^+$$

where $a^+ = \text{Max}(a,0)$.

Table 1. Percentage of negative estimates of σ^2 for normal populations

Sets	I		II		III		IV		V		VI		VII		VIIII	
	V[a]	%	V	%	V	%	V	%	V	%	V	%	V	%	V	%
Stage 1	1	24.0	4	6.0	9	1.8	9	7.7	9	16.6	9	18.8	9	21.9	1	46.0
2	1	18.1	1	18.1	1	18.1	4	2.8	9	0.6	9	4.5	9	12.6	9	12.6
3	1	3.9	1	3.9	1	3.9	1	3.9	1	3.9	4	0.05	9	0.00	9	0.00

[a]V = Variance ratios

S_1 = Sum of squares for "within" groups.

S_2 = Sum of squares for "Between" groups.

k = Number of groups; n = Number of observations in each group.

2. Zacks estimator (43):

$$\hat{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{S_2}{(k+1)} - \frac{S_1}{k(n-1)}\right]^+ .$$

3. Tiao and Tan: σ^2_α component of the mode of the posterior distribution of $p(\mu,\sigma^2,\sigma^2_\alpha/y)$, with prior $d\mu \frac{d\sigma}{\sigma}\frac{d\tau}{\tau}$, $\tau^2 = \sigma^2 + n\sigma^2_\alpha$:

$$\hat{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{S_2}{(k+2)} - \frac{S_1}{k(n-1)+2}\right]^+ .$$

4. Tiao and Tan: σ^2_α component of the mode of the posterior distribution of $p(\frac{S_1}{\sigma^2}, \frac{2n\sigma^2_\alpha}{S_2}/y)$ with prior $d\mu \frac{d\sigma}{\sigma}\frac{d\tau}{\tau}$, $\tau^2 = \sigma^2 + n\sigma^2_\alpha$:

$$\hat{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{S_2}{(k+1)} - \frac{S_1}{k(n-1)-2}\right]^+ .$$

5. Stone and Springer: σ^2_α component of the mode of the posterior distribution of $p(\mu,\sigma^2,\sigma^2_\alpha/y)$ with prior $d\mu \frac{d\sigma}{\tau}\frac{d\tau}{\tau}$, $\tau^2 = \sigma^2 + n\sigma^2_\alpha$:

$$\hat{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{S_2}{k+3} - \frac{S_1}{k(n-1)+1}\right]^+ .$$

6. Tiao and Tan: Posterior mean of σ^2_α with prior $d\mu \frac{d\sigma}{\sigma}\frac{d\tau}{\tau}$, $\tau^2 = \sigma^2 + n\sigma^2_\alpha$:

$$\hat{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{S_2}{(k-3)} \cdot \frac{H_\phi\{\frac{(k-3)}{2},\frac{k(n-1)}{2}\}}{H_\phi\{\frac{(k-1)}{2},\frac{k(n-1)}{2}\}} - \frac{S_1}{k(n-1)-2} \cdot \frac{H_\phi\{\frac{(k-1)}{2},\frac{k(n-1)-2}{2}\}}{H_\phi\{\frac{(k-1)}{2},\frac{k(n-1)}{2}\}}\right] .$$

7. Stone and Springer: posterior mean of σ_α^2 ; prior $d\mu \frac{d\sigma}{\tau}\frac{d\tau}{\tau}$; $\tau^2 = \sigma^2 + n\sigma_\alpha^2$:

$$\hat{\sigma}_\alpha^2 = \frac{1}{n}\left[\frac{S_2}{(k-2)} \cdot \frac{H_\phi\{\frac{(k-2)}{2}, \frac{k(n-1)-1}{2}\}}{H_\phi\{\frac{k}{2}, \frac{k(n-1)-1}{2}\}} - \frac{S_1}{k(n-1)-3}\, \frac{H_\phi\{\frac{k}{2}, \frac{k(n-1)\ 3}{2}\}}{H_\phi\{\frac{k}{2}, \frac{k(n-1)-1}{2}\}}\right]$$

where

$$\phi = \frac{S_2}{S_1} \quad \text{and} \quad H_\phi(a,b) = \frac{1}{\beta(a,b)} \int_0^{\phi/(1+\phi)} x^{a-1}(1-x)^{b-1}dx \quad .$$

They have tabulated mean square errors and mean values of the above mentioned estimators for $\sigma^2 + \sigma_\alpha^2 = 1$ and $\rho = \frac{\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2} = 0(.1)1.0$ and (k,n) pairs: (4,2), (5,2), (4,4), (4,8), (4,10), (10,4), (10,10), (5,50), (50,10) and (50,50). The results indicate that Bayes posterior means, considered here at 6, 7 above, have large mean square error which is due to infinite Bayes risks. The maximum likelihood estimator has larger mean square error than that of Bayes posterior modes. The Zacks estimator compares favourably with Bayes posterior mode for large values of ρ .

Besides the Zacks' estimator, considered above, Zacks (43) has proposed three other estimators of σ_α^2 . He uses the following quantities in his estimators:

$$D(R,n,k) = \frac{k(n-1)+2}{(n-1)^2(k-2)} \cdot \frac{1 - I_r[\frac{k}{2} - 2, \frac{k(n-1)}{2} + 2]}{1 - I_r[\frac{k}{2} - 1, \frac{k(n-1)}{2} + 1]}$$

$$E(R,n,k) = \frac{(n-1)^2(n-2)}{k(n-1)+2} \cdot \frac{1 - I_r[\frac{k}{2}, \frac{k(n-1)}{2}]}{1 - I_r[\frac{k}{2}-1, \frac{k(n-1)}{2}+1]}$$

where $r = [1 + (n-1)^2R]^{-1}$, $R = (S_2+kn\bar{Y}^2_{..})/S_1$ and $I_r(a,b)$ is the cumulative incomplete Beta with parameters a,b .

The estimators of σ_α^2 proposed by Zacks are as under:

i) $\hat{\sigma}_{\alpha_1}^2 = \frac{1}{n}[\frac{S_2}{k+1} - \hat{\sigma}_1^2]^+$

where

$$\hat{\sigma}_1^2 = \text{Min}[\frac{S_1}{k(n-1)+2}, \frac{S_1+S_2}{kn+1}, \frac{S_1+S_2+kn\bar{Y}_{..}^2}{kn+2}]$$

ii) $\hat{\sigma}_{\alpha_2}^2 = \frac{1}{n}\{\frac{S_1}{k+1} - \hat{\sigma}_2^2]^+$

where

$$\hat{\sigma}_2^2 = \frac{S_1}{(kn+2)}[1+D(R,n,k)] \ , \ R = \frac{S_2+kn\bar{Y}_{..}^2}{S_1} \qquad \text{if} \quad \mu = 0$$

and

$$\hat{\sigma}_2^2 = \frac{S_1}{(kn+1)}[1+D(R',n,k-1)] \ , \ R' = S_2/S_1 \quad \text{if} \quad \mu \neq 0 \ .$$

Here $\hat{\sigma}_2^2$ is an estimator of σ^2, constructed by taking the best estimator of σ^2 when $\rho = \frac{\sigma_\alpha^2}{\sigma^2}$ is known. This will be a function of ρ. Then a Bayes estimator of ρ is substituted in the function to yield $\hat{\sigma}_2^2$.

iii) $\hat{\sigma}_{\alpha_3}^2 = \frac{1}{n}[\frac{S_2}{k+1} - \frac{S_1}{k(n-1)}]^+$.

This is the estimator considered by Klotz and Milton (23):

iv) $\hat{\sigma}_{\alpha_4}^2$ is constructed in the same way as $\hat{\sigma}_2^2$.

If $\mu = 0$, then

$$\hat{\sigma}_{\alpha_4}^2 = \frac{S_2+kn\bar{Y}_{..}^2}{n(kn+2)}[\{E(R,n,k) + 1\} - \frac{1}{R}\{D(R,n,k) + 1\}] \ ,$$

where

$$R = \frac{S_2+kn\bar{Y}_{..}^2}{S_1} \ .$$

If $\mu \neq 0$, then

$$\hat{\sigma}^2_{\alpha_4} = \frac{S_2}{n(kn+1)}[\{E(R',n,k-1)+1\} - \frac{1}{R}.\{D(R',n,k-1)+1\}]$$

where

$$R' = S_2/S_1 \quad .$$

Zacks has conducted Monte Carlo studies of his estimators. Using mean square error as a criteria, the relative efficiency of the unbiassed estimator as compared with $\sigma^2_{\alpha 1}$, $\sigma^2_{\alpha 2}$, and $\sigma^2_{\alpha 3}$ is very low especially when $\rho = 0$. The efficiency functions of $\hat{\sigma}^2_{\alpha i}$, $(i = 1,2,3)$ are very close. All the three estimators slightly over estimate σ^2_{α} when $\rho = o$ as they never assume negative values. When $\rho > 0$, these estimators under-estimate σ^2_{α}, this being due to the divisor of S_2. The comparative study of $\hat{\sigma}^2_{\alpha 3}$ and $\hat{\sigma}^2_{\alpha 4}$ reveals that the differences in mean square errors are not striking. The estimators become more efficient as the number of groups (k) increases.

B. Approximate Confidence Intervals

The works of Bartlett (3,4), Bulmer (9), Huitson (21), Green (15), Williams (41), and Welch (39) in this field are note-worthy. Using a normal approximation, Huitson (21) and Welch (39) describe methods of assigning a confidence interval to a linear combination of variances of the form $V = \Sigma\lambda_i\sigma^2_i$, where λ_i are arbitrary constants and σ^2_i are such that the sample estimates S^2_i are independently distributed as $\sigma^2_i\chi^2_i/f_i$ with f_i degrees of freedom. Obviously the method is applicable to the one-way random model as $n\sigma^2_{\alpha} = (\sigma^2+n\sigma^2_{\alpha}) - \sigma^2$, and the sample estimates of $(\sigma^2+n\sigma^2_{\alpha})$, and σ^2 have the desired property. Welch (39) makes a distinction between the cases (i) λ_i are all positive and (ii) λ_i have different signs. The approximation is of the order $\frac{1}{f}$ and for large values of f_i, it is reasonably good. Huitson (21) gives a confidence interval of $\mu = \sum_i \lambda_i S_i^2 / \sum_i \lambda_i \sigma_i^2$. Writing $V_{mn} = \sum_i \frac{\lambda_i^m S_i^{2m}}{f_i^n} / (\sum_i \lambda_i S_i^2)^m$, the Pth percentage point of μ is

$$1 + \xi\sqrt{2V_{21}} + \xi^2[2V_{21} - \frac{4V_{32}}{3V_{21}}] - \frac{2}{3}\frac{V_{32}}{V_{21}} \quad \text{to order } \frac{1}{f}$$

where ξ is the P percent standard normal deviate. He has tabulated lower and upper 5 percent and 1 percent critical values of $\frac{\lambda_1 S_1^2 + \lambda_2 S_2^2}{\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2}$ for f_1, $f_2 = 16, 36, 144$ and ∞ ; $\frac{\lambda_1 S_1^2}{\lambda_1 S_1^2 + \lambda_2 S_2^2} = 0.0(0.1)1.0$.

When all λ_i are positive, Welch (39) has approximated the distribution of $\frac{v}{V} = \frac{\sum_i \lambda_i S_i^2}{\sum_i \lambda_i \sigma_i^2}$ by the standard χ_F^2 distribution with F degrees of freedom where

$$F = (\sum_i \lambda_i \sigma_i^2)^2 / \sum_i \lambda_i^2 \sigma_i^4 f_i^{-1}$$

$$\doteqdot (\sum_i \lambda_i S_i^2)^2 / \sum_i \lambda_i^2 S_i^4 f_i^{-1} .$$

More exhaustive investigations of this case have been carried out by Satterthwaite (29) and Box (6). Welch (39) has obtained approximate confidence intervals for (i) v-V and (ii) $\frac{v}{V}$ when λ_i have different signs.

He suggests that the use of $\frac{v}{V}$ is appropriate when V is essentially positive (as in the case of $n\sigma_\alpha^2$). Writing $r_{ab} = \sum_i \lambda_i^a S_i^{2a} f_i^{-b}$, the Pth percentage points are:

i) for $v - V$

$$\xi(2r_{12})^{\frac{1}{2}} - \frac{2}{3}(2\xi^2+1)r_{32}(r_{12})^{-1} \quad \text{to order } \frac{1}{f}$$

where ξ is the P percent standard normal deviate,

ii) for v/V

$$M_F - 2(2\xi^2+1)F^{-1}(r_{32}r_{21}^2 v-1)/3 \quad \text{to order } \frac{1}{f}$$

where

$$F = (\sum_i \lambda_i S_i^2)^2 \Big/ \sum_i (\lambda_i^2 S_i^4 f_i^{-1}) .$$

M_F = P percent critical value of standard χ^2 with F degrees of freedom and ξ = P percent standard normal deviate.

It may be noted that the above methods may yield a negative limit if $S_1 < S_2$ and we wish to find a confidence limit of $\sigma_1 - \sigma_2$. In such a case the limit has to be taken as zero if $\sigma_1^2 - \sigma_2^2$ is essentially positive.

The solution to approximate confidence limits given by Green (15) is very complicated and not explicit. It is incapable, therefore, of application unless it has been tabulated and no tables have been prepared so far. It has an advantage, in one sense, that it does not assume all the f_i large; in case of the one-way model he assumes f_2, the number of degrees of freedom for the within mean square, to be large, but not necessarily f_1, the number of degrees of freedom for the between mean square.

Bartlett (3,4) obtains confidence intervals from the log-likelihood derivatives using the normal approximation for one unknown parameter θ and for two parameters θ_1, and θ_2. Assuming $\frac{\partial L}{\partial \theta}$ as normal variate, a first approximation to a confidence interval for θ is given by

$$\frac{\partial L}{\partial \theta} = \pm \mu \sqrt{I(\theta)}$$

where μ is the appropriate normal deviate for desired confidence level and $I(\theta)$ is Fisher's information. The second approximate confidence interval for θ, correcting for skewness of $\frac{\partial L}{\partial \theta}$, is

$$\frac{\partial L}{\partial \theta} - \frac{1}{6} \frac{K_3}{I^2(\theta)} [(\frac{\partial L}{\partial \theta})^2 - I(\theta)] = \pm \mu \sqrt{I(\theta)}$$

where K_3 is the third cumulant of $\frac{\partial L}{\partial \theta}$.

For two parameters θ_1, θ_2, he assumes that the joint distribution of $\frac{\partial L}{\partial \theta_1}$, $\frac{\partial L}{\partial \theta_2}$ is approximately normal with zero means and known co-variance matrix. The confidence region for (θ_1, θ_2) is

$$(\frac{\partial L}{\partial \theta_1})^2 / I_{11} + (\frac{\partial L}{\partial \theta_2} - \frac{I_{12}}{I_{11}} \frac{\partial L}{\partial \theta_1})^2 / (I_{12} - \frac{I_{12}^2}{I_{11}}) \leq \chi_o^2$$

where $I_{ij} = E[-\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}]$ and χ_o^2 is the critical value of χ^2 for two degrees of freedom. He suggests that a correction for skewness is hardly necessary but outlines a procedure for such correction.

The case of two parameters θ_1, θ_2 with θ_2 as nuisance parameter is complicated. The approximate confidence interval is obtained through a series of stages. The standardized variable

$$T = \left(\frac{\partial L}{\partial \theta_1} - \frac{I_{12}}{I_{22}}\frac{\partial L}{\partial \theta_2}\right) / \sqrt{I_{11 \cdot 2}}$$

where $I_{11 \cdot 2} = I_{11} - I_{12}^2/I_{22}$ is presumed as a normal deviate with zero mean and unit variance. For the nuisance parameter θ_2 appearing in T, the best estimate of θ_2, given θ_1, is substituted. Finally correction is made for the skewness of T. The method is illustrated by considering the Fisher (13) fiducial probability argument in the variance component problem, where in the notation of Bartlett, we have χ^2 - quantities

$$\frac{nv}{\theta_1 + \lambda\theta_2}, \frac{n_2 v_2}{\theta_2}$$

with n and n_2 degrees of freedom of the between and the within mean squares respectively and we require a confidence interval of θ_1 .

The confidence interval for θ_1 is given by the solution of either of the two equations correct to $O(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n_2}})$:

$$T - \frac{1}{6}\gamma_1(T^2-1) = \pm \mu$$

$$T - \frac{1}{6}\gamma_1(\mu^2-1) = \pm \mu$$

where

μ = normal deviate for desired confidence,

$$T = (v-\lambda v_2-\theta_1)\sqrt{\left[\frac{nn_2}{2\{n_2(\theta_1+\lambda\theta_2)^2+\lambda^2 n\theta_2^2\}}\right]} \quad ,$$

$$\gamma_1 = 8\left\{\frac{(\theta_1+\lambda\theta_2)^3}{n_2} - \frac{\lambda^3\theta_2^3}{n_2^2}\right\} I_{11 \cdot 2}^{3/2}$$

and for θ_2 appearing in T and γ_1, we substitute

$$\hat{\theta}_2 = [\frac{n_2}{v_2} + \frac{\lambda n(v-\theta_1)}{(\theta_1+\lambda v_2)^2}]/[\frac{n_2}{v_2} + \frac{\lambda^2 n}{(\theta_1+\lambda v_2)^2}] \quad .$$

The utility of the results for data sets of the size commonly met is unknown.

Williams (41) constructs a confidence interval of σ^2_α by combining two intervals about functions of the parameters of interest and nuisance parameters. Let S_1 and S_2 denote sum of squares "within" and "between" groups with $k(n-1)$ and $(k-1)$ degrees of freedom respectively. The usual $(1-\alpha)$ percent confidence interval for the function $\sigma^2 + n\sigma^2_\alpha$ is

$$\frac{S_2}{\chi^2_{\alpha\mu}} \leq \sigma^2 + n\sigma^2_\alpha \leq \frac{S_2}{\chi^2_{\alpha\ell}}$$

or

$$\frac{1}{n}[\frac{S_2}{\chi^2_{\alpha u}} - \sigma^2] \leq \sigma^2_\alpha \leq \frac{1}{n}[\frac{S_2}{\chi^2_{\alpha\ell}} - \sigma^2]$$

where $\chi^2_{\alpha\mu}$ and $\chi^2_{\alpha\ell}$ are upper and lower χ^2 limits enclosing $(1-\alpha)$ percent of the distribution with $(k-1)$ degrees of freedom. Call this interval $I_1(\sigma^2)$.

Similarly a $(1-\alpha)$ percent confidence interval for $\frac{\sigma^2_\alpha}{\sigma^2}$, obtained by considering a confidence interval of $\sigma^2/(\sigma^2+n\sigma^2_\alpha)$, is

$$[\frac{k(n-1)S_2/(k-1)S_1 - F_{\alpha\mu}}{nF_{\alpha\mu}}]\sigma^2 \leq \sigma^2_\alpha \leq \sigma^2[\frac{k(n-1)S_2/(k-1)S_1 - F_{\alpha\ell}}{nF_{\alpha\ell}}]$$

where $F_{\alpha u}$ and $F_{\alpha \ell}$ are upper and lower $F(n_1,n_2)$ limits enclosing $(1-\alpha)$ percent distribution. Call this interval $I_2(\sigma^2)$. Williams has shown that for a given σ^2

$$(1-2\alpha) \leq P[\sigma^2_\alpha \subset I_1(\sigma^2) \ , \ \sigma^2_\alpha \subset I_2(\sigma^2)] < 1 - \alpha \ .$$

As the lower and upper bounds are independent of σ^2 , the probability statement holds good whatever may be the value of nuisance parameter σ^2 . The intersection of $I_1(\sigma^2)$ and $I_2(\sigma^2)$ when projected onto the σ^2_α axis is bounded by the intersection of the upper limits of the intervals and the intersection of lower limits of intervals. Thus, the interval

$$\frac{1}{n\chi^2_{\alpha u}}[S_2 - \frac{(k-1)S_1F_{\alpha u}}{k(n-1)}] \leq \sigma^2_\alpha \leq \frac{1}{n\chi^2_{\alpha \ell}} [S_2 - \frac{(k-1)S_1F_{\alpha \ell}}{k(n-1)}]$$

is true with a frequency greater than $(1-2\alpha)$ percent regardless of the true value of σ^2 .

Bulmer (9) adopts a different approach to solve the problem of approximate confidence interval of σ^2_α . Let M_1 and M_2 be independent mean square variates with f_1 and f_2 degrees of freedom and unknown expectations $(\theta+\sigma^2)$ and σ^2 respectively where θ and σ^2 are non-negative. Here θ corresponds to our $n\sigma^2_\alpha$. The problem is to find a function of M_1 and M_2 , $f(M_1,M_2)$ such that

$$Pr[f(M_1,M_2) \leq \theta] = \alpha$$

whatever may be θ and σ^2 .

As any change in the scale results in a corresponding change in the confidence limits, we must have $f(c^2M_1,c^2M_2) = c^2 f(M_1,M_2)$ for any real constant c . Letting $c^2 = 1/M_2$, we can write $f(M_1,M_2)$ in the form

$M_2 g(F)$ where $F = M_1/M_2$. He places the following limiting conditions on $g(F)$:

i) $g(L_1) = 0$,

ii) $g(F) \sim F/L_2$ as $F \to \infty$,

where

L_1 = Lower 100 α percent F-value with f_1 and f_2 degrees of freedom,

L_2 = Lower 100 α percent F-value with f_1 and ∞ degrees of freedom.

The approximate confidence limit, so obtained by Bulmer is

$$M_2[F/L_2 - 1 + (L_1/F)(1 - L_1/L_2) \quad \text{to the order} \quad (f_1/f_2)^2 \ .$$

It may be noted that for lower 100 α percent confidence limit of θ , F-values for L_1 and L_2 are taken at 100α percent cumulative point of F-distribution (with appropriate degrees of freedom) and for upper 100 α percent confidence limit of θ , F-values are taken at $100(1-\alpha)$ percent cummulative point.

The confidence limits are exact if $\rho = \frac{\theta}{\sigma^2} = 0$ or $\rho \to \infty$. They are also exact for all values of ρ when f_2 is infinite (f_1 remaining finite). For intermediate cases, the accuracy depends on the ratio of f_1 to f_2 and not on particular values of f_1 and f_2 .

C. Fiducial Limits

The concept of fiducial inference is due to Fisher (13) and the following remarks are thought to represent briefly Fisher's views. The type of argument is not generally accepted as being valid. The fiducial form of argument is purported to be a rigorous probability statement about the unknown parameter, which may be a vector, of the population without the assumption of any a priori knowledge about the probability distribution of the parameter. It should be distinguished from the inverse probability statement which seeks to specify the frequency with which the parameter would lie in an assigned range. The resulting statement is claimed to be true of the aggregate of cases in which the observed sample yields particular values of sample statistics, used in deriving the probability distribution. The fiducial probability statement is independent of any prior knowledge and is true of the aggregate of all samples. The sample statistics are the parameters of the distribution. Fisher states that in general, if statistics $T_1, T_2, T_3, \ldots$ contain jointly the whole of the information available about the parameters $\theta_1, \theta_2, \theta_3, \ldots$ and if functions $t_1, t_2, t_3, \ldots$ of the T's and θ's can be found, the simultaneous distribution of which is independent of $\theta_1, \theta_2, \theta_3, \ldots$, then the fiducial distribution of $\theta_1, \theta_2, \ldots$ may be found by substitution. If sufficient statistics exists, then the T's should be functions of sufficient statistics. The use of any other statistics for the fiducial distribution of θ's is improper as this amounts to the utilization of a part of the information and throwing away the rest. Consider, for example, the one-way random-effect model. A symbolic representation of the analysis of variance is:

Source	d.f.	M.S.	E.M.S.
Between	$n_1 = k-1$	V_1	$\sigma_1^2 = \sigma^2 + n\sigma_\alpha^2$
Within	$n_2 = k(n-1)$	V_2	$\sigma_2^2 = \sigma^2$.

The statistics (V_1, V_2) jointly are said to contain all available information about the parameters $(\sigma^2, \sigma_\alpha^2)$. Now let X_1, X_2 be random variables distributed independently as chi-square with n_1 and n_2 degrees of freedom respectively.

Then

$$V_1 = \frac{X_1 \sigma_1^2}{n_1} , \; V_2 = \frac{X_2 \sigma_2^2}{n_2} \quad .$$

The general procedure is to note that

$$\frac{n_1 V_1}{\sigma_1^2} \sim X_1 \quad \text{and} \quad \frac{n_2 V_2}{\sigma_2^2} \sim X_2 \quad .$$

The fiducial argument uses the typical inversion

$$P\{\frac{n_1 V_1}{\sigma_1^2} \leq k_1\} = F_{n_1}(k_1)$$

where $F_{n_1}(k_1)$ is the value of the cumulative distribution of a chi-square variate with n_1 degrees of freedom at any real number k_1 .

So

$$P(\sigma_1^2 \geq \frac{n_1 V_1}{k_1}) = F_{n_1}(k_1) \quad .$$

Let

$$\ell_1 = \frac{n_1 V_1}{k_1} , \; k_1 = \frac{n_1 V_1}{\ell_1}$$

then

$$P(\sigma_1^2 \geq \ell_1) = F_{n_1}(\frac{n_1 V_1}{\ell_1})$$

and

$$P(\sigma_1^2 \geq \ell_1 + \delta\ell_1) = F_{n_1}(\frac{n_1 V_1}{\ell_1 + \delta\ell_1})$$

$$= F_{n_1}(\frac{n_1 V_1}{\ell_1} - \frac{n_1 V_1 \delta\ell_1}{\ell_1^2}) \quad .$$

So

$$P(\ell_1 \leq \sigma^2 \leq \ell_1 + \delta\ell_1) = F_{n_1}(\frac{n_1V_1}{\ell_1}) - F_{n_1}(\frac{n_1V_1}{\ell_1} - \frac{n_1V_1\delta\ell_1}{\ell_1^2})$$

$$= \frac{n_1V_1}{\ell_1^2}\,\delta\ell_1 f_{n_1}(\frac{n_1V_1}{\ell_1}) \quad .$$

Replacing ℓ_1 by σ_1^2 , we have the fiducial density of σ_1^2 as

$$f_F(\sigma_1^2)d\sigma_1^2 = f_{n_1}(\frac{n_1V_1}{\sigma_1^2})\frac{n_1V_1}{(\sigma_1^2)^2}d\sigma_1^2$$

where F indicates that the density is fiducial.

Similarly we have the fiducial density of σ_2^2 as

$$f_F(\sigma_2^2)d\sigma_2^2 = f_{n_2}(\frac{n_2V_2}{\sigma_2^2})\,\frac{n_2V_2}{(\sigma_2^2)^2}\,d\sigma_2^2$$

and finally the joint fiducial density of (σ_1^2, σ_2^2) is

$$f_F(\sigma_1^2,\sigma_2^2)d\sigma_1^2 d\sigma_2^2 = f_F(\sigma_1^2)d\sigma_1^2 f_F(\sigma_2^2)d\sigma_2^2 \quad .$$

According to Fisher we can investigate this to get the fiducial density of $\sigma_1^2 - \sigma_2^2$ which equals to $n\bar{\sigma}_\alpha^2$. We note that there has been much controversy about this whole process and the process has been rejected by many as being based on logical misapprehensions.

The tabulation of the above defined distribution of $(\sigma_1^2-\sigma_2^2)$ presents difficulties of integration. Bross (8) has given approximate fiducial limits of

$$L = \frac{\frac{n_1V_1}{X_1} - \frac{n_2V_2}{X_2}}{V_1 - V_2} = \frac{F\frac{n_1}{X_1} - \frac{n_2}{X_2}}{F-1}$$

where $F = \frac{V_1}{V_2}$. Now

$$P(L \leq L_o) = P[\frac{1}{F-1}(\frac{n_1 F}{X_1} - \frac{n_2}{X_2}) \leq L_o]$$

$$= P[X_1 \geq \frac{n_1 F}{(F-1)L_o + \frac{n_2}{X_2}}]$$

$$= P[X_1 \geq z]$$

where

$$z = \frac{n_1 F}{(F-1)L_o + \frac{n_2}{X_2}} .$$

So

$$P(L \leq L_o) = \int_0^\infty [f(x_2) \int_z^\infty f(x_1) dx_1] dx_2 .$$

where $f(x_1)$ is the density of X_1 and $f(x_2)$ is the density of X_2.

Direct integration is difficult as z is a function of X_2. The following limiting cases admit a direct solution for $P(L \leq L_o) = \alpha$:

(i) $n_2 \to \infty$.

Then

$$\frac{n_2}{X_2} \to 1 \qquad \lim_{n_2 \to \infty} z = \frac{n_1 F}{(F-1)L_o + 1}$$

and the solution is

$$L_o = \frac{F/F'_\alpha - 1}{F-1}$$

where F'_α is the entry in the F-table for d.f. (n_1, ∞) .

(ii) $F \to \infty$.

Then

$$\lim_{F \to \infty} z = \frac{n_1}{L_o}$$

and the solution is $L_o = 1/F'_\alpha$.

(iii) $F = F_\alpha$.

Then

$$L_o = 0 \text{ is the solution.}$$

Bross considers the following function for an approximate fiducial interval,

$$L = \frac{(F/F_\alpha)-1}{(FF'_\alpha/F_\alpha)-1} \qquad F \geq F_\alpha$$

$$= 0 \qquad F < F_\alpha$$

which is exact in limitng cases considered above as

$$\lim_{n_2 \to \infty} L = \frac{(F/F'_\alpha)-1}{(FF'_\alpha/F'_\alpha)-1} = \frac{(F/F'_\alpha)-1}{F-1}$$

$$\lim_{F \to \infty} L = \frac{1}{F'_\alpha} \quad \text{and} \quad \lim_{F \to F_\alpha} L = 0 \ .$$

The approximate (1-2α) percent fiducial limits which are in close agreement with exact fiducial limits are

$$\underline{L} = \frac{(F/F_\alpha)-1}{(FF'_\alpha/F_\alpha)-1} \qquad F \geq F_\alpha$$

$$= 0 \qquad F < F_\alpha$$

$$\bar{L} = \frac{(F/F_{(1-\alpha)})-1}{(FF'_{(1-\alpha)}/F_{(1-\alpha)})-1}$$

where

F is the ratio V_1/V_2 obtained from the data

F_α is the entry in a α percent F-table for n_1, n_2 d.f.

F'_α is the entry in a α percent F-table n_1, ∞ d.f.

Healy (17) works with the fiducial distribution of

$$\phi = \frac{n\sigma_\alpha^2}{V_1} = \frac{n_1}{X_1} - \frac{n_2}{FX_2} \quad .$$

$$P_r[\phi > \phi_o] = \int_R f(X_1)f(X_2)dX_1dX_2 .$$

where R is the region defined by

$$(X_1 - \frac{n}{\phi_o})(X_2 + \frac{n_2}{F\phi_o}) + \frac{n_1 n_2}{F\phi_o^2} < 0 \quad .$$

This can be written as:

$$P_r[\phi > \phi_o] = \int_o^{\frac{n}{\phi_o}}[f(X_1) \int_y^\infty f(X_2)dX_2]dX_1$$

where

$$y = \frac{n_2X_1/n_1}{F(1-\phi_o X_1/n_1)} \quad .$$

The distribution of ϕ is complicated, but can be evaluated without difficulty of both n_1 and n_2 are even. Healy has tabulated the distribution function $\phi = n\sigma_\alpha^2/V_1$ for $n_1, n_2 = 6, 8, 12, 24$ and ∞ ; $F = 0.5$, 1.0, 2.0, 4.0, 8.0, 16.0 and ∞ and $P = .95$, $.05$, $.99$ and $.01$. The tabulated values have been published and appear in (14). A serious limitation is that the exact fiducial limits for values of n_1 and n_2 other than even are not available. The use is, therefore, restricted to the cases where both n_1 and n_2 are even or in other words the number of groups as well as the number of observations in each group in the sample are odd.

D. Bayesian Approach

The essential difference between non-Bayesian inference and Bayesian inference is the form of the question that we ask ourselves and try to answer. In non-Bayesian inference, we try to answer the question; what does a sample have to say about the parameter(s) denoted by θ ? The answer may be in the form of a point estimate of θ , α percent confidence interval of θ , a fiducial distribution of θ , testing of some hypothesis about θ, etc. No prior knowledge of θ is assumed except its space, denoted by Ω . The inference is based on the sample alone. Any prior knowledge of the subject-matter is presumed to have been utilized in the planning of the experiment. In Bayesian inference, it is assumed that every experimenter has a prior knowledge H, called prior beliefs, about θ . It is assumed, further, that this prior knowledge, however vague it may be, can be expressed in the form of a distribution of θ , at least locally. After a sample has been observed, the Bayesian inference results in a change in beliefs about θ , to what are called posterior beliefs. According to this approach, the accumulation of knowledge is a continuous process. For the purpose of the next sample, the posterior belief becomes the prior belief which changes into a new posterior belief, after another sample has been observed. A Bayesian process gives, in the limit, perfect knowledge about the parameter, and thus is thought to be logically compelling. We note in passing, however, that all procedures seem to have this convergence property.

Suppose that an experimenter has a sample $X = (x_1, x_2, \ldots, x_n)$ of n observations $x_i (i=1 \ldots n)$ having identical and independent distribution defined by the density $f(x_i/\theta)$, where θ is the fixed but unknown parameter(s). Let H denote his state of knowledge about θ before he has observed the sample. Then θ will have a distribution $p(\theta/H)$ depending on H. This prior distribution of θ will represent his prior degree of beliefs about θ. The density of the random sample $X = (x_1, x_2, \ldots, x_n)$ will be

$$p(X/\theta,H) = \prod_{i=1}^{n} f(x_i/\theta) \;.$$

The density of his beliefs about θ will be changed by the sample from $p(\theta/H)$ to $p(\theta/X,H)$, called the posterior density of θ. According to Bayes's Theorem,

$$p(\theta/X,H) \propto p(X/\theta,H)p(\theta/H) \quad .$$

Thus, posterior density of $\theta \propto$ Likelihood of Sample X prior density of θ.

At the next stage in the progress of knowledge about θ, $p(\theta/X,H)$, the posterior density of θ, will be the prior density of θ which will change into a new posterior denstiy of θ when the experimenter observes another random sample. Thus, the terms, the prior density of θ and the posterior density of θ are relative to the state of knowledge of the experimenter in a process which will lead to a perfect knowledge of θ in the limit.

Tiao, Tan and Box (34,35,36) and Hill (19,20) have recently produced solutions of the problem from the Bayesian point of view in a series of papers. Hill (20) has developed a "general" theory of inference from a one-way random model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad (i=1, \ldots ,k; j=1, \ldots n)$$

$\alpha_i \sim N(0,\sigma^2_\alpha)$, $\epsilon_{ij} \sim NID(0,\sigma^2)$. α_i and ϵ_{ij} are mutually independent. He uses what is termed a diffuse or "non-informative" prior, namely,

$$p(\sigma^2,\sigma^2_\alpha)\alpha(\sigma^2)^{-1}(\sigma^2+n\sigma^2_\alpha)^{-1} \qquad \sigma^2 \geq 0 \ , \ \sigma^2_\alpha \geq 0$$

which assumes that log σ^2 and $\log(\sigma^2+n\sigma^2_\alpha)$ have locally uniform prior distributions.

Tiac and Tan (35) using the same non-informative prior present posterior distributions for

$$w = 1 + n\sigma^2_\alpha/\sigma^2$$

$$\tau = S_1/\sigma^2$$

$$\mu = 2n\sigma^2_\alpha/S_2 \quad \text{and} \quad (\tau,u)$$

where $S_1 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij}-\bar{y}_{i.})^2$ = sum of squares within groups

$S_2 = n \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$ = sum of squares between groups.

The posterior distributions obtained by them are

(i) posterior distribution of $w = 1 + n\sigma^2_\alpha/\sigma^2$

$$p(w/Y) = \frac{1}{H_\phi\{\frac{1}{2}(k-1),\frac{1}{2}k(n-1)\}} \; \frac{\phi^{-\frac{1}{2}k(n-1)}}{\beta\{\frac{1}{2}(k-1),\frac{1}{2}k(n-1)\}} \; w^{\frac{1}{2}k(n-1)-1}\left(1+\frac{w}{\phi}\right)^{-\frac{1}{2}(kn-1)}$$

$$1<w<\infty$$

where $\phi = S_2/S_1$

$$H_\phi(a,b) = \frac{1}{\beta(a,b)}\int_0^{\phi/(1+\phi)} x^{a-1}(1-x)^{b-1}dx \quad .$$

The distribution has the form of a truncated F distribution

(ii) joint posterior distribution of $\tau = S_1/\sigma^2$, $u = 2n\sigma_\alpha^2/S_2$

$$p(\tau,u/Y) = D\tau^{\frac{1}{2}k(n-1)-1}\left(\frac{\phi\tau}{1+\frac{1}{2}\phi\tau u}\right)^{\frac{1}{2}(k-1)+1}\mathrm{Exp}\left[-\tfrac{1}{2}\left\{\tau + \frac{\phi\tau}{1+\frac{1}{2}\phi\tau u}\right\}\right]$$

where

$$D^{-1} = \Gamma\{\tfrac{1}{2}(k-1)\}\ \Gamma\{\tfrac{1}{2}k(n-1)\}\ H_\phi\{\tfrac{1}{2}(k-1),\tfrac{1}{2}k(n-1\}2^{\frac{1}{2}(kn+1)} \quad .$$

The distribution has a unique mode.

$$\left.\begin{aligned}\tau_o &= k(n-1) - 2\\ \mu_o &= \frac{2}{k+1} - \frac{1}{\phi}\left(\frac{2}{k(n-1)-2}\right)\end{aligned}\right\} \quad \text{if} \quad \phi \geq \frac{k+1}{k(n-1)-2}$$

$$\left.\begin{aligned}\tau_o &= \frac{k(n-1)}{1+\phi}\\ \mu_o &= 0\end{aligned}\right\} \quad \text{if} \quad \phi < \frac{k+1}{k(n-1)-2} \quad .$$

(iii) Marginal posterior distribution of $\tau = S_1/\sigma^2$

$$p(\tau/Y) = f_\tau\{k(n-1)\}\cdot\frac{G_{\frac{1}{2}\phi\tau}\{\frac{1}{2}(k-1)\}}{H_\phi\{\frac{1}{2}(k-1),\frac{1}{2}k(n-1)\}}$$

where $f_\tau\{k(n-1)\}$ is $\chi^2_{k(n-1)}$ and $G_w(p)$ is the incomplete Gamma integral

$$G_w(p) = \frac{1}{\Gamma(p)}\int_0^w x^{p-1}e^{-x}dx \quad .$$

This gives

$$E(\tau^r/Y) = 2^r \frac{\Gamma\{\frac{1}{2}k(n-1)+r\}}{\Gamma\{\frac{1}{2}k(n-1)\}} \cdot \frac{H_\phi\{\frac{1}{2}(k-1),\frac{1}{2}k(n-1)+r\}}{H_\phi\{\frac{1}{2}(k-1),\frac{1}{2}k(n-1)\}}$$

and the moment generating functions is

$$M_\tau(t) = \frac{H_{\phi/(1-2t)}\{\frac{1}{2}(k-1),\frac{1}{2}k(n-1)\}}{H_\phi\{\frac{1}{2}(k-1),\frac{1}{2}k(n-1)\}} (1-2t)^{-\frac{1}{2}k(n-1)}, |t| < \tfrac{1}{2} \quad .$$

For large values of $k(n-1)$, the distribution of τ can be well approximated by a scaled $a\chi_b^2$ where $a > 0$ is a scalar multiplier and b represents the degrees of freedom. The parameters a and b are obtained by equating the first two moments of τ to that of $a\chi_b^2$.

(iv) The marginal posterior distribution of $\mu = 2n\sigma_\alpha^2/S_2$

$$p(\mu/Y) = 2^{\frac{1}{2}(kn+1)} \cdot D \int_0^\infty h(\mu,z) z^{\frac{1}{2}k(n-1)-1} e^{-z} dz$$

where D is as given in (ii) and

$$h(\mu,z) = [(\phi z)^{-1}+\mu]^{-\frac{1}{2}(k+1)} \mathrm{Exp}\{-[(\phi z)^{-1}+\mu]^{-1}\} \quad .$$

The distribution is defined over the range $(0,\infty)$. The mode lies in the interval.

$$0 \leq \mu \leq 2[\frac{1}{k-1} - \frac{1}{\phi(kn-3)}]$$

if $\phi > \frac{k+1}{kn+3}$, and is at the origin if otherwise.

According to the Bayesian analysis, the problem of a negative estimate of σ_α^2, which we face in the sampling theory approach, does not exist.

The posterior distribution of σ_α^2 is defined over the range $(0,\infty)$ and the mode can, at most, be at the origin which corresponds to the case of a negative (sample theory) estimate of σ_α^2. In case the posterior mean is taken as the estimate, it is always positive.

Stone and Springer (31) suggest the following non-informative prior

$$d\mu \frac{d\sigma}{\bar{\sigma}} \cdot \frac{d\bar{\sigma}}{\bar{\sigma}}$$

where

$$\bar{\sigma}^2 = \sigma^2 + n\sigma_\alpha^2 \ .$$

Faced with a negative (traditional) estimate of σ_α^2 or the mode of posterior distribution of $2n\sigma_\alpha^2/S_2$ at the origin in the Bayesian analysis, one may suspect the inadequacy of the model. B. M. Hill (19) considers the following model admitting within-class correlation.

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad i=1, \ldots ,I; \quad j=1, \ldots ,J$$

$$\alpha_i \sim N(0,\sigma_\alpha^2) \ ; \quad \epsilon_{ij} \sim N(0,\sigma^2)$$

$$E(\alpha_i \epsilon_{i'j}) = 0 \quad \text{for all} \quad i,i' \quad \text{and} \quad j$$

$$E(\epsilon_{ij}\epsilon_{i'j'}) = 0 \quad \text{for} \quad i \neq i' \quad \text{and} \quad j,j'$$

$$E(\epsilon_{ij}\epsilon_{ij'}) = C \quad \text{for} \quad j \neq j' \ .$$

The likelihood function is

$$L(\mu,\sigma^2,\sigma_\alpha^2, C/Y)\alpha(\sigma^2-C)^{-I(J-1)/2}[(\sigma^2-C)+J(\sigma_\alpha^2+C)]^{-I/2}X$$

$$\text{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2-C} + \frac{S_2+IJ(\mu-\bar{Y}..)^2}{(\sigma^2-C)+J(\sigma_\alpha^2+C)}]\}$$

$- \infty < \mu < \infty$; $\sigma^2 \geq 0$; $\sigma_\alpha^2 \geq 0$; $-\sigma^2/(J-1) < C \leq \sigma^2$.

It is obvious that the likelihood function depends on $(\mu, \sigma^2 - C, \sigma_\alpha^2 + C)$. The fundamental unidentifiability of the problem is clear. Hill (19) has shown that under quite weak conditions on the prior distribution of μ , $\sigma_\alpha^2, \nu^2 = \sigma^2/S_1$, $y = 1 - \frac{C}{\sigma^2}$, the joint posterior distributon of $(\mu, \sigma_\alpha^2, \nu^2, y)$ converges to a limiting distribution as S_1 goes to ∞ . There is no unique way to characterize vague prior knowledge. He uses the following convenient prior

$$p(\mu, \sigma^2, \sigma_\alpha^2, y) \alpha (\sigma^2)^{-1} [\sigma_\alpha^2 + \sigma^2(1-y/y_o)]^{-1} f_\alpha(\sigma_\alpha^2) \quad .$$

where $f_\alpha(\sigma_\alpha^2)$ has a finite integral over bounded sets and $y_o = J/(J-1)$.

The limiting posterior distribution of σ_α^2 , obtained by him is

$$p'(\sigma_\alpha^2) \; \alpha \; f_\alpha(\sigma_\alpha^2) \Pr[\chi^2_{(I-1)} \leq \frac{S_2}{J\sigma_\alpha^2}] \quad .$$

Tiao and Tan (36) have considered the effect of auto-correlated errors. The model is

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad (i=1, \ldots k \; ; \; j=1, \ldots n)$$

$$e_{ij} = \rho e_{i(j-1)} + \epsilon_{ij} \; , \; -\infty < \rho < \infty$$

$$\alpha_i \sim NID(0, \sigma_\alpha^2) \; ; \; \epsilon_{ij} \sim NID(0, \sigma^2) \quad .$$

The following transformation is made

$$z_{i_o} = y_{i_1} = \mu + \alpha_i + \rho e_{i_o} + \epsilon_{i_1}$$

$$z_{i_t} = y_{i_{(t+1)}} - \rho y_{i_t} = (1-\rho)\mu + (1-\rho)\alpha_i + \epsilon_{i_{(t+1)}}$$

$$(i=1, \ldots k \; ; \; t=1, \ldots m \; ; \; m=n-1) \quad .$$

Using the following prior distribution, given a value of ρ :

$$p((1-\rho)\mu,(1-\rho)^2\sigma^2_\alpha,\sigma^2/\rho)\alpha(\sigma^2)^{-1}[\sigma^2+m(1-\rho^2)\sigma^2_\alpha]^{-1} .$$

They have obtained the following posterior distributions

(i) <u>The joint posterior distribution of $V = \frac{1}{\sigma^2}$, σ^2_α</u>

$$p(V,\sigma^2_\alpha/\rho,Y)\alpha(1-\rho)^2 V^{[\frac{1}{2}(km+1)-1]}[1+m(1-\rho)^2\sigma^2_\alpha V]^{-[\frac{1}{2}(k-1)+1]}\times$$

$$\mathrm{Exp}\{-\frac{V}{2}[S_1(\rho)+\frac{S_2(\rho)}{1+m(1-\rho)^2\sigma^2_\alpha V}]\} \qquad (V>0\ ,\ \sigma^2_\alpha>0)$$

where

$$S_1(\rho) = \sum_{i=1}^{k}\sum_{t=1}^{m}(Z_{i_t}-\bar{Z}_i)^2 ,$$

$$S_2(\rho) = m\sum_{i=1}^{k}(\bar{Z}_i-\bar{Z})^2 .$$

Let $\phi(\rho) = S_2(\rho)/S_1(\rho)$. The distribution has a unique mode:

$$V_o = \frac{k(m-1)-2}{S_1(\rho)} ,$$

$$\sigma^2_{\alpha_o} = \frac{1}{m(1-\rho)^2}[\frac{S_2(\rho)}{k+1}-\frac{S_1(\rho)}{k(m-1)-2}] \quad \text{if} \quad \phi(\rho) > \frac{k+1}{k(m-1)-2}$$

otherwise,

$$V_o = \frac{km-1}{S_1(\rho)+S_2(\rho)} ; \ \sigma^2_{\alpha_o} = 0 .$$

(ii) <u>The marginal posterior distribution of $V = \frac{1}{\sigma^2}$</u>

$$p(V/\rho,Y)\alpha S_1(\rho)f_T\{k(m-1)\}\cdot\frac{G_{\frac{1}{2}\phi(\rho)T}\{\frac{1}{2}(k-1)\}}{H_{\phi(\rho)}\{\frac{1}{2}(k-1),\frac{1}{2}k(m-1)\}}$$

where $\tau = S_1(\rho)/\sigma^2$ and $G_{w(p)}$ and $H_{\phi(\rho)}(a,b)$ are as defined earlier. The distribution admits all moments.

(iii) The marginal posterior distribution of σ^2_α

$$p(\sigma^2_\alpha/\rho,Y) = Q(\rho)\int_0^\infty h(\mu,z)z^{[\frac{1}{2}k(m-1)-1]}e^{-z}dz$$

where

$$h(\mu,z) = \{[\phi(\rho)z]^{-1}+\mu\}^{-\frac{1}{2}(k+1)}\mathrm{Exp}\{-[(\phi(\rho)z)^{-1}+\mu]^{-1}\}$$

$$Q(\rho) = \frac{2m(1-\rho)^2}{S_2(\rho)\ \Gamma\{\frac{1}{2}(k-1)\}\ \Gamma\{\frac{1}{2}k(m-1)\}H_{\phi(\rho)}\{\frac{1}{2}(k-1),\frac{1}{2}k(m-1)\}} .$$

The distribution has a mode in the interval

$$0 < \sigma^2_\alpha < \frac{1}{m(1-\rho^2)}\left[\frac{S_2(\rho)}{k+1} - \frac{S_1(\rho)}{km+3}\right] \quad \text{if} \quad \phi(\rho) > \frac{k+1}{km+3}$$

other wise the mode is to be taken at the origin.

For $\frac{1}{2}(k-1) > r$, the r'th moment of the distribution exists.

The question of negative estimates of σ^2 , σ^2_α , therefore, does not exist according to the Bayesian point of view. The estimate of σ^2_α can at most be zero if we take the posterior mode as our estimate.

Recently Tiao and Box (34), using non-informative priors, have presented a Bayesian solution of three component hierarchical design model

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk} \qquad (i=1, \ldots I,\ J=1, \ldots J,\ k=1, \ldots K)$$

$$\alpha_i \sim N(0,\sigma_3^2),\ \beta_{ij} \sim NID(0,\sigma_2^2),\ \epsilon_{ijk} \sim NID(0,\sigma_1^2) .$$

A Bayesian estimation of means of one-way random effect model

$$y_{jk} = \theta_j + \epsilon_{jk} \qquad (j=1, \ldots J, k=1, \ldots K)$$

$$\theta_j \sim N(0,\sigma_2^2) \ \epsilon_{jk} \sim NID(0,\sigma_1^2)$$

has been presented by Box and Tiao (7) and results compared with fixed effect model.

A serious objection that can be raised against the Bayesian analysis is the use of non-informative prior. The final results depend upon the prior we use. A non-informative prior has no meaning. It is selected perhaps, for the ease of integration and to avoid negative estimate in an artificial way. If one has a prior knowledge of subject matter, however vague, it seems reasonable that the knowledge should be expressed in a mathematical form and that prior used.

III. BAYESIAN INFERENCE

In this chapter we consider Bayesian inference about certain parametric functions in the random-effect model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad (i=1, \ldots k\ , \ j=1, \ldots n)$$

$\alpha_i \sim NID(0,\sigma^2_\alpha)$, $\epsilon_{ij} \sim NID(0,\sigma^2)$, α_i and ϵ_{ij} independent.

The following notations will be used throughout this chapter.

$$\bar{Y}_{i.} = \sum_{j=1}^{n} Y_{ij} \Big/ n\ ,\ \bar{Y}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n} Y_{ij} \Big/ kn\ ,$$

$$S_1 = \sum_{i=1}^{k} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i.})^2 = \text{"Within" sum of squares,}$$

$$S_2 = n \sum_{i=1}^{k} (\bar{Y}_{i.} - \bar{Y}_{..})^2 = \text{"Between" sum of squares,}$$

$\nu_1 = k(n-1)$ = degrees of freedom for "within" mean square,

$\nu_2 = k-1$ = degrees of freedom for "between" mean square.

Tiao and Tan (35), using the non-informative prior

$$\frac{d\sigma^2}{\sigma^2} \frac{d(\sigma^2 + n\sigma^2_\alpha)}{(\sigma^2 + n\sigma^2_\alpha)}\ ,$$

have considered properties of posterior distributions of

$$\frac{S_1}{\sigma^2}\ , \frac{n\sigma^2_\alpha}{S_2} \quad \text{and} \quad 1 + \frac{\eta\sigma^2_\alpha}{\sigma^2}\ .$$

The non-informative prior used by Tiao and Tan (35) contains the element n which is a sample variable. It appears that they have used this prior for the ease of integration. As the likelihood function depends on σ^2 and $\sigma^2+n\sigma^2_\alpha$, this non-informative prior is conveniently absorbed in the likelihood function. We feel that a prior should be independent of the sample size and the ease of integration should not be an allurement to the use of a prior dependent on the sample size. We shall examine the use of prior distributions which are independent of the sample size. Our attention will be directed predominantly to the nature of the posterior mode.

A. Posterior Distribution of $(\sigma^2+\sigma^2_\alpha)/\sigma^2$ with Tan-Tiao-Like Prior

This represents the ratio of gross variance to the error variance. A posterior distribution gives information on this ratio. We can also draw an inference about σ^2_α/σ^2 from such a distribution. Let us consider the following non-informative prior similar to the Tan and Tiao prior.

$$d\mu \, \frac{d\sigma^2}{\sigma^2} \cdot \frac{d\tau^2}{\tau^2} \quad \text{where} \quad \tau^2 = \sigma^2 + \sigma^2_\alpha \ .$$

The likelihood function is

$$L(\mu,\sigma^2,\sigma^2_\alpha/Y)\alpha(\sigma^2)^{-\nu_1/2}(\sigma^2+n\sigma^2_\alpha)^{-(\nu_2+1)/2}X$$

$$\text{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2} + \frac{S_2}{\sigma^2+n\sigma^2_\alpha} + \frac{nk(\bar{Y}_{..} - \mu)^2}{\sigma^2+n\sigma^2_\alpha}]\}$$

$$\sigma^2 \geq 0 \ ; \ \sigma^2_\alpha \geq 0 \ ; \ -\infty < \mu < \infty$$

and the joint posterior distribution of (μ,σ^2,τ^2) is

$$p(\mu,\sigma^2,\tau^2/Y)\alpha(\sigma^2)^{-(\nu_1/2+1)}(\tau^2)^{-1}\{n\tau^2-(n-1)\sigma^2\}^{-(\nu_2+1)/2}\text{X}$$

$$\text{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2}+\frac{S_2}{\{n\tau^2-(n-1)\sigma^2\}}+\frac{nk(\bar{Y}..-\mu)^2}{\{n\tau^2-(n-1)\sigma^2\}}]\}d\mu d\sigma^2 d\tau^2$$

$$-\infty<\mu\ \infty\ ,\ \sigma^2\geq 0\ ,\ \sigma^2\leq\tau^2\leq\infty\ .$$

After integrating μ out, we have the posterior distribution of (σ^2,τ^2)

$$p(\sigma^2,\tau^2/Y)\alpha(\sigma^2)^{-(\nu_1/2+1)}(\tau^2)^{-1}\{n\tau^2-(n-1)\sigma^2\}^{-\nu_2/2}\text{X}$$

$$\text{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2}+\frac{S_2}{n\tau^2-(n-1)\sigma^2}]\}d\sigma^2 d\tau^2\ .$$

$$\sigma^2\geq 0,\sigma^2\leq\tau^2\leq\infty$$

Making the transformation

$$w=\sigma^2,\ z=\frac{\tau^2}{\sigma^2}=\frac{\sigma^2+\sigma^2_\alpha}{\sigma^2}\ ,$$

we have

$$p(w,z/Y)\alpha(z)^{-1}w^{-\{(\nu_1+\nu_2)/2+1\}}\{nz-(n-1)\}^{-\nu_2/2}\text{X}$$

$$\text{Exp}\{-\frac{1}{2w}[S_1+\frac{S_2}{nz-(n-1)}]\}dwdz$$

$$w\geq 0,\ 1\leq z\leq\infty\ .$$

Integrating w out, we have the marginal posterior distribution of z.

$$p(z/Y) = C \cdot \frac{[nz-(n-1)]^{\nu_1/2}}{z[\{nz-(n-1)\}S_1+S_2]^{(\nu_1+\nu_2)/2}} \qquad 1 \leq z \leq \infty$$

where C is the normalizing constant. The integration for the normalizing constant is not straightforward, but it can be easily done by numerical analysis methods.

1. Moments of the posterior distribution: $(n-1)S_1 \neq S_2$

$$E(z^r/Y) = C \cdot \int_1^\infty \frac{z^{r-1}[nz-(n-1)]^{\nu_1/2}}{[\{nz-(n-1)\}S_1+S_2]^{(\nu_1+\nu_2)/2}}\,dz \quad .$$

Let $x = nz - (n-1)$, then

$$E(z^r/Y) = C \cdot \int_1^\infty \frac{[x+(n-1)]^{r-1}.x^{\nu_1/2}}{n^r(xS_1+S_2)^{(\nu_1+\nu_2)/2}}dx$$

$$= \frac{C}{n^r}\int_1^\infty [\sum_{m=o}^{r-1}\binom{r-1}{m}(n-1)^{r-m-1}\frac{x^{\nu_1/2+m}}{(xS_1+S_2)^{(\nu_1+\nu_2)/2}}]dx$$

$$= \frac{C}{n^r}\sum_{m=o}^{r-1}[\binom{r-1}{m}(n-1)^{r-m-1}\int_1^\infty \frac{x^{\nu_1/2+m}}{(xS_1+S_2)^{(\nu_1+\nu_2)/2}}\,dx] \quad .$$

Let

$$I = \int_1^\infty \frac{x^{\nu_1/2+m}}{(xS_1+S_2)^{(\nu_1+\nu_2)/2}}dx \; .$$

Making the transformation $y = x/\phi$, where $\phi = S_2/S_1$, we have

$$I = \frac{1}{S_1^{\nu_1/2+(m+1)}S_2^{\nu_2/2-(m+1)}}\int_{1/\phi}^\infty \frac{y^{\nu_1/2+m}}{(1+y)^{(\nu_1+\nu_2)/2}}\,dy \quad .$$

Letting $y = \frac{1-\mu}{\mu}$,

$$I = \frac{1}{S_1^{\nu_1/2+(m+1)} S_2^{\nu_2/2-(m+1)}} \cdot \int_0^{\phi/(1+\phi)} \mu^{\frac{\nu_2}{2} - (m+2)} (1-\mu)^{\frac{\nu_1}{2} + m} \, d\mu$$

$$= \frac{1}{S_1^{\nu_1/2+(m+1)} S_2^{\nu_2/2-(m+1)}} \beta\left(\frac{\nu_2}{2} -(m+1), \frac{\nu_1}{2}+(m+1)\right) H_\phi\left(\frac{\nu_2}{2}-(m+1), \frac{\nu_1}{2}+(m+1)\right)$$

where $H_\phi(a,b)$ is the incomplete beta integral

$$H_\phi(a,b) = \frac{1}{\beta(a,b)} \int_0^{\frac{\phi}{(1+\phi)}} x^{a-1}(1-x)^{b-1} dx \quad .$$

Finally

$$E(z^r/y) = \frac{C}{n^r} \sum_{m=0}^{r-1} \binom{r-1}{m} (n-1)^{r-m-1} \frac{\beta\left(\frac{\nu_2}{2}-(m+1), \frac{\nu_1}{2}+(m+1)\right) H_\phi\left(\frac{\nu_2}{2}-(m+1), \frac{\nu_1}{2}+(m+1)\right)}{S_1^{\nu_1/2+(m+1)} S_2^{\nu_2/2-(m+1)}} \quad .$$

For $r \leq \frac{k-1}{2}$, the rth moment exists and the posterior mean is

$$E(z/y) = \frac{C}{n} \cdot \frac{\beta\left(\frac{\nu_2}{2}-1, \frac{\nu_1}{2}+1\right) H_\phi\left(\frac{\nu_2}{2}-1, \frac{\nu_1}{2}+1\right)}{S_1^{\nu_1/2+1} \, S_2^{\nu_2/2-1}} \quad .$$

2. Mode of the posterior distribution : $(n-1)S_1 \neq S_2$

Using $\nu_1 = k(n-1)$, $\nu_2 = k-1$,

$$\frac{\partial p(z/y)}{\partial z} = \frac{C}{2} \cdot \frac{[nz-(n-1)]^{\frac{k(n-1)}{2} - 1}}{z^2 [nS_1 z - \{(n-1)S_1 - S_2\}]^{\frac{kn+1}{2}}} \times Q(z)$$

where

$$Q(z) = kn(n-1)z[nS_1 z-\{(n-1)S_1-S_2\}]-n(kn-1)S_1 z[nz-(n-1)]$$
$$-2[nz-(n-1)][nS_1 z-\{(n-1)S_1-S_2\}] \quad .$$

Collecting coefficients of the quadratic $Q(z)$, we have

$$\begin{aligned}
z^2 &: -n^2(k+1)S_1 && < 0 \\
z &: n(n-1)(k+3)S_1 + n[k(n-1)-2]S_2 && > 0 \\
\text{const} &: -2(n-1)[(n-1)S_1 - S_2] \ . && \lesseqgtr 0 \ \text{ if } \ (n-1)S_1 \gtrless S_2
\end{aligned}$$

The discriminant function D of the quadratic is

$$\begin{aligned}
D &= [n(n-1)(k+3)S_1 + n[k(n-1)-2]S_2]^2 - 8n^2(n-1)(k+1)S_1[(n-1)S_1 - S_2] \\
&= n^2(n-1)^2(k-1)^2S_1^2 + n^2[k(n-1)-2]^2S_2^2 \\
&\quad + 2n^2(n-1)[(k+3)\{k(n-1)-2\} + 4(k+1)]S_1S_2 > 0 \quad \text{as} \quad k(n-1) \geq 2 \ .
\end{aligned}$$

Hence the quadratic has two real and distinct roots for all permissible values of k, n, S_1 and S_2. The roots have opposite sign if $(n-1)S_1 < S_2$ and are both positive if $(n-1)S_1 > S_2$.

<u>Case (i)</u> $(n-1)S_1 < S_2$. $\frac{\partial p(z/y)}{\partial z}$ can be written as

$$\frac{\partial p(z/y)}{\partial z} = -\psi(z)A(z+m_1)(z-m_2) \qquad 0 < m_1 < m_2$$

where

$$\psi(z) = \frac{C}{2}\,\frac{[nz-(n-1)]^{\frac{k(n-1)}{2}-1}}{z^2[nS_1z-\{(n-1)S_1-S_2\}]^{\frac{kn+1}{2}}} \qquad > 0$$

for $1 \leq z < \infty$.

$$A = n^2(k+1)S_1 > 0$$

and $-m_1$ and m_2 are the roots of the quadratic $Q(z)$.

We are interested in m_2 only which may lie within or outside the range.

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_2} = -\psi(m_2)A(m_1+m_2) \quad < 0$$

so that m_2 is a point of maxima. The posterior distribution will have a mode if $m_2 > 1$, otherwise the mode will be at $Z_0 = 1$.

<u>Case (ii)</u> $(n-1)S_1 > S_2$. The quadratic has two positive real roots. Suppose that the distinct roots are m_1 and m_2 satisfying $0 < m_1 < m_2$, we will show that at most m_2 can lie between 1 and ∞ and $p(z/y)$ is either uni-modal or has no mode at all in the range over which it is defined.

Consider $p(z/y)$ as a function of z, say $f(z)$. Obviously, $f(z)$ can be extended over the range $1 - \frac{1}{n} \leq z \leq \infty$. Then $f(z)$ is a continuous, non-negative valued function, nowhere zero except at the ends i.e. $f(\frac{n-1}{n}) = f(\infty) = 0$. Hence, $f(z)$ has a relative maxima between $\frac{n-1}{n}$ and ∞. Now

$$\frac{\partial f(z)}{\partial z} = -\psi(z)A(z-m_1)(z-m_2) \qquad 0 < m_1 < m_2$$

where A and $\psi(z)$ are as defined under Case (i). Moreover $\psi(z) > 0$, for $\frac{n-1}{n} < z < \infty$ holds. Differentiating again and evaluating the second derivative at m_1 and m_2 we have

$$\left.\frac{\partial^2 f(z)}{\partial z^2}\right|_{m_1} = -\psi(m_1)A(m_1-m_2) \qquad > 0$$

$$\left.\frac{\partial^2 f(z)}{\partial z^2}\right|_{m_2} = -\psi(m_2)A(m_2-m_1) \qquad < 0$$

Suppose that $\frac{n-1}{n} < m_1 < m_2$ holds. Then m_1 is the point of relative minima and m_2 the point of relative maxima. As $f(z)$ is a non-negative continuous function with $f(\frac{n-1}{n}) = f(\infty) = 0$, $f(z)$ can not be monotone decreasing function between $\frac{n-1}{n}$ and m_1. Hence, m_1 must lie in the interval $(0, \frac{n-1}{n}>$. Of course $m_1 = \frac{n-1}{n}$ can hold. The existence of mode of the posterior distribution depends on m_2. If $m_2 > 1$, the posterior distribution has a mode. If $\frac{n-1}{n} \leq m_2 \leq 1$, then mode is to be taken at $Z_0 = 1$.

Finally for $(n-1)S_1 \neq S_2$, $p(z/y)$ is either uni-modal or has no mode in the range $1 \leq z \leq \infty$. In no case is $p(z/y)$ bi-modal.

3. <u>Mode of the posterior distribution $(n-1)S_1 = S_2$</u>

The posterior density $p(z/y)$, has the simple form

$$p(z/y) \alpha \frac{[nz-(n-1)]^{\frac{k(n-1)}{2}}}{z^{(kn+1)/2}}$$

and

$$\frac{\partial p(z/y)}{\partial z} \alpha \frac{[nz-(n-1)]^{\frac{k(n-1)}{2}-1}}{z^{(kn+1)/2+1}} \left[\frac{(n-1)(kn+1)}{n(k+1)} - z\right] .$$

Hence the mode is given by

$$z_o = \frac{(n-1)(kn+1)}{n(k+1)} > 0$$

and $z_o > 1$ if $kn(n-2) > 1$. Thus, the posterior distribution of z has a mode in the range $(1,\infty)$ if $n \geq 3$. If $n = 2$, then $z_o = \frac{2k+1}{2(k+1)} < 1$, however large k may be and, therefore, no mode exists for $n = 2$. The Bayesian estimate of $(\sigma^2+\sigma_\alpha^2)|\sigma^2$ is Γ for all values of k.

B. Posterior Distribution of $(\sigma^2+\sigma_\alpha^2)/\sigma^2$ with Stone-Springer-Like Prior

Next we consider the following Stone-Springer-like prior.

$$d\mu \frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau} \quad \sigma \geq 0 \ , \ \tau^2 = \sigma^2 + \sigma_\alpha^2 \ , \ \sigma \leq \tau \leq \infty \ .$$

The joint posterior distribution of (σ,τ) obtained by integrating out μ from the joint posterior distribution of (μ,σ,τ) is

$$p(\sigma,\tau/y)\alpha \frac{1}{\tau^2}(\sigma^2)^{-\nu_1/2}[n\tau^2-(n-1)\sigma^2]^{-\nu_2/2} \mathrm{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2} + \frac{S_2}{n\tau^2-(n-1)\sigma^2}]\} \ .$$

$$\sigma \geq 0, \ \sigma \leq \tau \leq \infty \qquad d\sigma d\tau$$

Making the transformation

$$w = \frac{1}{\sigma^2} \ , \ z = \frac{\tau^2}{\sigma^2} = \frac{\sigma^2+\sigma_\alpha^2}{\sigma^2}$$

we have the joint posterior distribution of (w,z)

$$p(w,z/y)\alpha(w)^{(\nu_1+\nu_2)/2-1} z^{-3/2}[nz-(n-1)]^{-\nu_2/2} \mathrm{Exp}\{-\frac{w}{2}[S_1 + \frac{S_2}{nz-(n-1)}]\} \ .$$

$$w \geq 0 \ , \ 1 \leq z \leq \infty \qquad dwdz$$

Integrating w out, we have the posterior distribution of z

$$p(z/y)\alpha \frac{[nz-(n-1)]^{\nu_1/2}}{z^{3/2}[nS_1z-\{(n-1)S_1-S_2\}]^{(\nu_1+\nu_2)/2}} dz \ .$$

$$1 \leq z \leq \infty$$

The integration for the normalizing constant and moments of distribution is not simple, but numerical methods are available for this purpose.

1. Mode of the posterior distribution; $(n-1)S_1 \neq S_2$

Differentiating $p(z/y)$ with respect to z and using $\nu_1 = k(n-1)$, $\nu_2 = k-1$ we have

$$\frac{\partial p(z/y)}{\partial z} \propto \frac{[nz-(n-1)]^{\frac{k(n-1)-2}{2}}}{z^{5/2}\{nS_1 z - [(n-1)S_1 - S_2]\}^{\frac{kn+1}{2}}} \times Q(z)$$

where

$$\begin{aligned} Q(z) = {} & kn(n-1)z\{nS_1 z-[(n-1)S_1-S_2]\} \\ & - 3[nz-(n-1)]\{nS_1 z-[(n-1)S_1-S_2]\} \\ & - n(kn-1)S_1 z[nz-(n-1)] \quad . \end{aligned}$$

Collecting the coefficients of the quadratic, we have

$$\begin{array}{lll} z^2: & -n^2(k+2)S_1 & < 0 \\ z: & n(n-1)(k+5)S_1+n[k(n-1)-3]S_2 & > 0 \text{ , } (k,n) \neq (2,2) \\ \text{Const:} & -3(n-1)[(n-1)S_1-S_2] & \lesseqgtr 0 \text{ if } (n-1)S_1 \gtrless S_2 \text{ .} \end{array}$$

Hence $Q(z)$ has one positive and one negative real root if $(n-1)S_1 < S_2$ and has two or none positive real roots if $(n-1)S_1 > S_2$.

Let us consider the discriminant function D .

$$\begin{aligned} D &= \{n(n-1)(k+5)S_1 + n[k(n-1)-3]S_2\}^2 - 12n^2(n-1)(k+2)[(n-1)S_1-S_2] \\ &= n^2(n-1)^2(k-1)^2S_1^2 + n^2[k(n-1)-3]^2S_2^2 \\ &\quad + 2n^2(n-1)[k^2(n-1) + k(5n-2)-3]S_1S_2 \\ &\quad > 0 \text{ for all permissible values of } k, n, S_1 \text{ and } S_2 \text{ .} \end{aligned}$$

Thus, $Q(z)$ has real and distinct roots. If $(n-1)S_1 < S_2$, then $Q(z)$ has real roots with opposite signs and if $(n-1)S_1 > S_2$ then $Q(z)$ has two real positive roots.

<u>Case (i)</u> $(n-1)S_1 < S_2$. We can write $\frac{\partial p(z/y)}{\partial z}$ in the following form.

$$\frac{\partial p(z/y)}{\partial z} \,\alpha\, - \psi(z).A.(z+m_1)(z-m_2)$$

where

$$\psi(z) = \frac{[nz-(n-1)]^{\frac{k(n-1)-2}{2}}}{z^{5/2}\{nS_1 z-[(n-1)S_1-S_2]\}^{\frac{kn+1}{2}}} > 0$$

for $\frac{(n-1)}{n} < z < \infty$.

$$A = n^2(k+2)S_1$$

and $(-m_1)$ and (m_2) are the real roots of $Q(z)$ satisfying $0 < m_1 < m_2$.

We are not interested in $z_o = -m_1$, being outside the range of $p(z/y)$. It is easy to verify that $z_o = m_2$ is the point of relative maxima, provided that $p(z/y)$ as a function of z is well defined for $z = m_2$. If $m_2 > 1$, then the mode of the posterior distribution lies in the interval $(1,\infty)$. If $m_2 \leq 1$, then there is no mode and the Bayesian estimate of $\frac{\sigma^2+\sigma_\alpha^2}{\sigma^2}$ is 1.

<u>Case (ii)</u> $(n-1)S_1 > S_2$. The quadratic $Q(z)$ has two positive and distinct real roots say, m_1 and m_2, where $0 < m_1 < m_2$. It may be noted that the posterior distribution of z, as a function of z, can be extended to cover the range $\frac{n-1}{n} \leq z \leq \infty$. Then $f(z)$ is a continuous, non-negative valued function which is nowhere zero except at $z_o = \frac{n-1}{n}$ and $z_o = \infty$ i.e. at both ends of the range over which it is

defined. We need only to repeat the arguments, previously advanced, to show that at most m_2 can lie in the interval $(1,\infty)$. The posterior distribution has a mode if $m_2 > 1$, otherwise the mode is to be taken at $z_o = 1$ and the Bayesian estimate of $(\sigma^2+\sigma_\alpha^2)/\sigma^2$ is 1.

Thus, we reach the same conclusion that for $(n-1)S_1 \neq S_2$, $p(z/y)$ is either uni-modal or has no mode depending upon the greater of the two roots of the quadratic $Q(z)$. In no case, the posterior distribution is bi-modal.

2. Mode of the posterior distribution, $(n-1)S_1 = S_2$

The posterior distribution has the simple form

$$p(z/y) \alpha \frac{[nz-(n-1)]^{\frac{k(n-1)}{2}}}{z^{kn/2+1}}$$

and

$$\frac{\partial p(z/y)}{\partial z} \alpha \frac{[nz-(n-1)]^{\frac{k(n-1)}{2}-1}}{z^{kn/2+2}} \cdot \left[\frac{(n-1)(kn+2)}{n(k+2)} - z\right].$$

Hence, the mode is given by

$$z_o = \frac{(n-1)(kn+2)}{n(k+2)} > 0 .$$

Now $z_o > 1$ if $kn(n-2) > 2$. Thus, the posterior distribution has the mode in the interval $(1,\infty)$ if $n \geq 3$. For $n = 2$, the mode is at $z_o = \frac{k+1}{k+2}$. The posterior distribution has no mode in the interval $(1,\infty)$ for $n = 2$, however large may be k and, therefore, the Bayesian estimate of $\frac{\sigma^2+\sigma_\alpha^2}{\sigma^2}$ is 1 for $n = 2$ and for all values of k.

C. The Posterior Distribution of $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$ with Tiao-Tan-Like Prior

In one-way random effect model, a parameter of great interest is the intraclass correlation $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$. We consider the posterior distribution of $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$, using a Tiao-Tan-like non-informative prior

$$d\mu \frac{d\sigma^2}{\sigma^2} \frac{d\tau^2}{\tau^2}$$

where $\tau^2 = \sigma^2 + \sigma_\alpha^2$; $\sigma^2 \geq 0$; $\sigma^2 \leq \tau^2 \leq \infty$; $-\infty < \mu < \infty$.

The joint posterior distribution of (σ^2, τ^2), obtained by integrating out μ from joint posterior distribution of (μ, σ^2, τ^2), is

$$p(\sigma^2, \tau^2/y) \propto (\tau^2)^{-1}(\sigma^2)^{-(\nu_2/2+1)}(\sigma^2+n\sigma_\alpha^2)^{-\nu_2/2}\text{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2} + \frac{S_2}{\sigma^2+n\sigma_\alpha^2}]\} .$$

Making the transformation

$$w = \frac{1}{\sigma^2} , \quad \frac{\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2} = z$$

we have

$$p(w,z/y) \propto \frac{w^{(\nu_1+\nu_2)/2-1}(1-z)^{\nu_2/2-1}}{[1+(n-1)z]^{\nu_2/2}} \text{Exp}\{-\frac{1}{2}w[S_1 + \frac{(1-z)S_2}{1+(n-1)z}]\}dwdz$$

$w \geq 0$, $0 \leq z \leq 1$.

Integrating w out, we have the marginal posterior distribution of z

$$p(z/y) \propto \frac{(1-z)^{\nu_2/2-1}[1+(n-1)z]^{\nu_1/2}}{[z\{(n-1)S_1-S_2\} + (S_1+S_2)]^{(\nu_1+\nu_2)/2}} \qquad 0 \leq z \leq 1 .$$

The integration for the normalizing constant C is not simple, generally, but it can be worked out by numerical methods. If both ν_1 and ν_2 are even i.e. the number of classes as well as the number of observations in each class are odd, then the calculation of the normalizing constant becomes simple. Let

$$\frac{\nu_1}{2} = k_1 \;, \; \frac{\nu_2}{2} = k_2 \;,$$

then by making the transformation

$$x = z[(n-1)S_1 - S_2] + (S_1+S_2)$$

the integrand reduces to the form

$$\sum_{i=1}^{k_1+k_2} b_i/x^i \quad \text{if} \quad (n-1)S_1-S_2 \neq 0$$

$$\sum_{i=o}^{k_1+k_2-1} a_i x^i \quad \text{if} \quad (n-1)S_1-S_2 = 0$$

where a_i and b_i are known constants. Thus, the normalizing constant is a sum of simple integrals, finite in number.

$$\frac{1}{C} = \begin{cases} \sum_{i=o}^{k_1+k_2-1} \int_o^1 a_i x^i dx & \text{if } nS_1 = S_1+S_2 \\ \sum_{i=1}^{k_1+k_2} \int_{nS_1}^{S_1+S_2} \frac{b_i}{x^i} dx & \text{if } nS_1 < S_1+S_2 \\ \sum_{i=1}^{k_1+k_2} \int_{S_1+S_2}^{nS_1} \frac{b_i}{x^i} dx & \text{if } nS_1 > S_1+S_2 \end{cases} .$$

Obviously all moments exist and can be calculated:

$$E(z^r/y) = \begin{cases} C \sum_{i=o}^{k_1+k_2-1} \int_o^1 a_i x^{i+r} dx & \text{if } nS_1 = S_1 + S_2 \\ C \sum_{i=1}^{k_1+k_2} \int_{nS_1}^{S_1+S_2} \frac{b_i}{x^{i-r}} dx & \text{if } nS_1 < S_1+S_2 \\ C \sum_{i=1}^{k_1+k_2} \int_{S_1+S_2}^{nS_1} \frac{b_i}{x^{i-r}} dx & \text{if } nS_1 > S_1+S_2 \end{cases}$$

1. Mode of the posterior distribution: $k > 4$, $(n-1)S_1 \neq S_2$

Writing ν_1 and ν_2 in terms of k, n,

$$p(z/y) \alpha \frac{[1+(n-1)z]^{\frac{k(n-1)}{2}} [1-z]^{\frac{k-3}{2}}}{[z\{(n-1)S_1-S_2\} + (S_1+S_2)]^{(kn-1)/2}} \qquad 0 \leq z \leq 1 \quad .$$

Let $(n-1)S_1-S_2 = a$, $S_1+S_2 = b$

$$\frac{\partial p(z/y)}{\partial z} = 0$$

gives

(i) $z = 1$, if $k \geq 5$ and at $z = 1$, $p(z/y) = 0$ which is the minimum for a density. It may be noted that for $0 \leq z < 1$, $p(z/y) \neq 0$.

(ii) $z = -\frac{1}{n-1}$ which is outside the range of $p(z/y)$ and, therefore, we are not interested.

(iii) roots of the following quadratic $Q(z)$ in z as possible maxima or minima.

$$Q(z) = k(n-1)^2(1-z)(za+b)-(k-3)(za+b)[1+(n-1)z]$$
$$- a(kn-1)(1-z)[1+(n-1)z] = 0 \quad .$$

Collecting terms for coefficients of the quadratic, $Q(z)$.

Coefficient of z^2

$$a(n-1)(kn-1)-a(n-1)(k-3)-ak(n-1)^2 = 2a(n-1) \lessgtr 0 \text{ if } a \lessgtr 0$$

$$= 2(n-1)^2S_1-2(n-1)S_2 \ .$$

Coefficient of z

$$k(n-1)^2(a-b)-(k-3)[a+b(n-1)]-a(kn-1)(n-2)$$

$$= a(n+1)-b(kn-3)(n-1)$$

$$= -(n-1)[n(k-1)-4]S_1-(kn^2-kn-2n+4)S_2 < 0 \text{ for } k \geq 4 \ .$$

Constant

$$bk(n-1)^2-b(k-3)-a(kn-1)$$

$$= b[kn(n-2)+3]-a(kn-1) > 0 \text{ if } a < 0$$

$$= S_2[kn(n-1)+2]-S_1[n(k-1)-2] \ .$$

Thus, the quadratic $Q(z)$ has the form

$$Q(z) = 2a(n-1)z^2 + [a(n+1) - b(kn-3)(n-1)]z$$
$$+ b[kn(n-2) + 3] - a(kn-1) \ .$$

For the existence of real roots of $Q(z)$, we have to examine the discriminant function D.

$$D = [a(n+1) - b(kn-3)(n-1)]^2$$
$$- 8a(n-1)[b(kn^2-2kn+3) - a(kn-1)]$$
$$= [(n+1)^2 + 8(n-1)(kn-1)]a^2 + (kn-3)^2(n-1)^2b^2$$
$$- 2(n-1)[(kn-3)(n+1) + 4(kn^2-2kn+3)]ab \ .$$

Expressing a, b in terms of S_1 and S_2 we have

$$D = (n-1)^2[(n+1)^2+8(n-1)(kn-1) + (kn-3)^2-2(kn-3)(n+1)$$
$$- 8(kn^2-2kn+3)]S_1^2 + [(n+1)^2+8(n-1)(kn-1) + (n-1)^2(kn-3)^2$$
$$+ 2(kn-3)^2 + 2(kn-3)(n^2-1) + 8(kn^2-2kn+3]S_2^2$$

$$+ 2(n-1)[(kn-3)^2(n-1) - (n+1)^2-8(n-1)(kn-1)$$
$$- (n-2)\{(kn-3)(n+1) + 4(kn^2-2kn+3)\}]S_1S_2 \quad .$$

After simplification, this is

$$n^2(n-1)^2(k-1)^2S_1^2 + n^2[k(n-1)+2]^2S_2^2$$
$$+ 2n^2(n-1)[k^2(n-1) - k(5n-3)+2]S_1S_2 \; .$$

Now

$$D = \{n(n-1)(k-1)S_1-n[k(n-1)+2]S_2\}^2$$
$$+ 2n^2(n-1)(k-1)[k(n-1)+2]S_1S_2$$
$$+ 2n^2(n-1)[k^2(n-1) - k(5n-3)+2]S_1S_2$$
$$= \{n(n-1)(k-1)S_1-n[k(n-1)+2]S_2\}^2$$
$$+ 4n^2k(n-1)^2(k-3)S_1S_2 \quad .$$

Hence, $D > 0$ for $k \geq 4$ and real and distinct roots exist for all permissible values of n, k, S_1 and S_2 .

It may be recalled that the coefficient of z in $Q(z)$ is negative for $k \geq 4$. We need consider the following cases.

<u>Case (i)</u> $a < 0$ i.e. $S_2/S_1 > (n-1)$. With $a < 0$, the constant in $Q(z)$ is positive and $Q(z)$ has the form $- Az^2 - Bz + C$; $A, B, C > 0$. By the rule of sign, two real roots with opposite signs exist. Now consider

$$Q(0) = C > 0 \; ,$$

$$Q(1) = S_1[2(n-1)^2-(n-1)\{n(k-1)-4\}- n(k-1)+2]$$
$$+ S_2[kn(n-1)+2-2(n-1) - (kn^2-kn-2n+4)]$$
$$= - n^2(k-3)S_1 < 0 \quad .$$

$Q(z)$ is a polynomial, with real coefficients, changing signs between 0 and 1. $Q(z)$, therefore, vanishes for some z_o , $0 < z_o$, < 1 . The positive real root lies in the interval (0,1), We can write $\frac{\partial p(z/y)}{\partial z}$ in the following form:

$$\frac{\partial p(z/y)}{\partial z} \alpha - A\psi(z)(z-m_1)(z+m_2)$$

$$0 < m_1 < m_2 \text{ , } A > 0$$

where m_1 and $-m_2$ are the roots of $Q(z)$ and

$$\psi(z) = \frac{[1+(n-1)z]^{\frac{k(n-1)-2}{2}} [1-z]^{\frac{k-5}{2}}}{[z\{(n-1)S_1-S_2\}+(S_1+S_2)]^{\frac{kn+1}{2}}} > 0 \quad \text{for} \quad 0 \leq z < 1 \quad .$$

Differentiating $p(z/y)$ with respect to z twice, we have

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_1} \alpha - A\psi(m_1)(m_1+m_2) \qquad < 0 \quad .$$

m_1 is, therefore, a maxima of $p(z/y)$. The posterior distribution of z has a mode in (0,1) and the Bayesian estimate

$$\left(\frac{\hat{\sigma}^2_\alpha}{\sigma^2+\sigma^2_\alpha}\right)$$

lies between (0,1) .

<u>Case (ii)</u> $a > 0$ and $S_2[kn(n-1)+2] < S_1[n(k-1)-2]$. $Q(z)$ has the form $Az^2 - Bz - C$. $A,B,C > 0$.

$\frac{\partial p(z/y)}{\partial z}$ can be written as

$$\frac{\partial p(z/y)}{\partial z} \alpha\ A\psi(z)(z-m_2)(z+m_1)$$

where $\psi(z)$ and A are as defined above and m_2, $-m_1$ are the roots of

$Q(z)$ satisfying $0 < m_1 < m_2$. Differentiating $p(z/y)$ with respect to z twice, we have

$$\frac{\partial^2 p(z/y)}{\partial z^2}\bigg|_{m_2} \alpha A\psi(m_2)(m_1+m_2) > 0 \quad .$$

If the positive real root of $Q(z)$ lies in $(0,1)$, then m_2 is a point of minima of $p(z/y)$ which is non-zero in $(0,1)$ and is zero at $z = 1$. A continuous density function can not have two consecutive minima. Hence, m_2 must lie outside $(0,1)$. At most $m_2 = 1$. The mode of the posterior density has, therefore, to be taken at the origin. Thus, the Bayesian estimate of

$$\left(\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha}\right)$$

is zero.

<u>Case (iii)</u> $a > 0, S_2[kn(n-1)+2] > S_1[n(k-1)-2]$. The condition means that S_1 and S_2 satisfy the relation

$$\frac{n(k-1)-2}{kn(n-1)+2} < \frac{S_2}{S_1} < (n-1) \quad .$$

$Q(z)$ has the form $Az^2 - Bz + C$ where $A,B,C > 0$. By the rule of sign, $Q(z)$ has none or two positive real roots, but no negative real roots. The possibility of no real roots is excluded by the consideration of the general case, where we have shown that real and distinct roots always exist. Suppose that m_1 and m_2 are positive real roots of $Q(z)$ and $0 < m_1 < m_2$. We will show that at most one positive real root can be in the interval $(0,1)$. Differentiating $p(z/y)$ with respect to z, we have

$$\frac{\partial p(z/y)}{\partial z} \alpha A\psi(z)(z-m_1)(z-m_2)$$

where $\psi(z)$ and A are as defined earlier.

Differentiating $p(z/y)$ with respect to z twice, we have

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_1} \alpha A\psi(m_1)(m_1-m_2) < 0 \quad ,$$

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_2} \alpha A\psi(m_2)(m_2-m_1) > 0 \quad .$$

It may be seen that m_1 and m_2 are points of maxima and minima respectively. We are faced with the same problem of two consecutive minima for a continuous density function and therefore m_2, must lie outside (0,1). At most $m_2 = 1$. It is possible that both m_1 and m_2 may be outside the range and we have no mode.

Finally for

$$\frac{n(k-1)-2}{kn(n-1)+2} < \frac{S_2}{S_1} < n-1$$

we have no definite conclusions about the existence of the mode of the distribution.

Special Case I $\quad (n-1)S_1 = S_2$, $k \geq 4$. The posterior distribution simplifies to

$$p(z/y)\alpha(1-z)^{\frac{k-3}{2}}[1+(n-1)z]^{\frac{k(n-1)}{2}} \quad .$$

Differentiating $p(z/y)$ with respect to z , we have

$$\frac{\partial p(z/y)}{\partial z} \alpha(1-z)^{\frac{k-5}{2}}[1+(n-1)z]^{\frac{k(n-1)-2}{2}}[(kn^2-2kn+3)-(n-1)(kn-3)z] \quad .$$

We have already discussed the points $z = 1$ and $z = -\frac{1}{n-1}$. The other possible point of maxima is

$$z_o = \frac{kn^2-2kn+3}{(n-1)(kn-3)} > 0$$

and $z_o < 1$ if $k > 3$ which is satisfied and we have the mode of $p(z/y)$ in the interval $(0,1)$.

Special Case II $\quad S_2/S_1 = [n(k-1)-2]/[kn(n-1)+2]$; $k \geq 4$.

$\frac{\partial p(z/y)}{\partial z}$ has the form $\psi(z)Az(z-B)$, $A,B > 0$, and differentiating $p(z/y)$ with respect to z again, we have

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{z=o} \alpha - AB\psi(0) < 0$$

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{z=B} \alpha AB\psi(B) > 0 \quad \text{if} \quad 0 < B < 1 \ .$$

It may be observed that $z = o$ is maxima and $z = B$ is minima. This, we have shown already, is not possible. B can at most be equal to 1. Thus, $p(z/y)$ has mode at zero and the Bayesian estimate of

$$\left(\frac{\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2}\right)$$

is zero.

We summarize the the results as follows:

(i) $\frac{S_2}{S_1} \geq n-1$, $p(z/y)$ has a mode in $(0,1)$.

(ii) $\frac{S_2}{S_1} \leq \frac{n(k-1)-2}{kn(n-1)+2}$, $p(z/y)$ has no mode in $(0,1)$ and therefore

$\left(\frac{\hat{\sigma}^2_\alpha}{\sigma^2+\sigma^2_\alpha}\right) = 0$.

(iii) $\frac{n(k-1)-2}{kn(n-1)+2} < \frac{S_2}{S_1} < n-1$

We can make no definite conclusion. The mode may or may not exist, depending on $n,k,\ S_1$ and s_2 .

(iv) In no case $p(z/y)$ is bi-modal.

D. The Posterior Distribution of $\sigma^2_\alpha/(\sigma^2+\sigma^2_\alpha)$ with Stone-Springer-Like Prior

We investigate the posterior distribution of $\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha}$ using a Stone-Springer-like prior $d\mu\,\frac{d\sigma}{\tau}\,\frac{d\tau}{\tau}$ where $\tau^2 = \sigma^2+\sigma^2_\alpha$, $\sigma^2 \geq 0$, $\sigma^2 \leq \tau^2 \leq \infty$, $-\infty \leq \mu \leq \infty$.

The joint posterior distribution of $(\sigma\ ,\tau\ )$ obtained by integrating out μ from joint posterior distribution of $(\mu,\sigma\ ,\tau\ )$ is

$$p(\sigma,\tau/y)\alpha(\sigma^2)^{-\nu_1/2}(\sigma^2+n\sigma^2_\alpha)^{-\nu_2/2}(\tau^2)^{-1}\mathrm{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2}+\frac{S_2}{\sigma^2+n\sigma^2_\alpha}]\}d\sigma d\tau \quad .$$

$$\sigma \geq 0, \sigma \leq \tau \leq \infty$$

Letting $w = \frac{1}{\sigma^2}$, $z = \frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha}$ we have

$$p(w,z/y)\alpha\,\frac{w^{\frac{(\nu_1+\nu_2)}{2}-1}(1-z)^{\frac{\nu_2-1}{2}}}{[1+(n-1)z]^{\nu_2/2}}\,\mathrm{Exp}\{-\tfrac{1}{2}\,w[S_1+\frac{(1-z)S_2}{1+(n-1)z}]\}dzdw \quad .$$

$$w \geq 0, 0 \leq z \leq 1$$

Integrating w out, we have the marginal posterior distribution of z.

$$p(z/y) \alpha \frac{[1+(n-1)z]^{\nu_1/2}(1-z)^{(\nu_2-1)/2}}{[z\{(n-1)S_1-S_2\}+(S_1+S_2)]^{(\nu_1+\nu_2)/2}} dz \quad .$$

$$0 \leq z \leq 1$$

The integration for the normalizing constant is complicated. It simplifies if ν_2 is odd and ν_1 is even i.e. k is even. It may be noted that for the Tio and Tan-like prior we require both k and n odd for the simplification of the integral and for the Stone-Springer-like prior, we have a restriction on k alone.

Making the transformation

$$x = z\{(n-1)S_1-S_2\} + (S_1+S_2)$$

the integrand is reduced to the form

$$\sum_{i=1}^{\nu_1+\nu_2} b_i/x^{i/2} \quad \text{if} \quad (n-1)S_1 \neq S_2 \quad ,$$

$$\sum_{i=0}^{(\nu_1+\nu_2-1)/2} a_i x^i \quad \text{if} \quad (n-1)S_1 = S_2$$

where a_i and b_i are known constants.

The normalizing constant is a sum of simple integrals, finite in number.

$$\frac{1}{C} = \begin{cases} \sum_{i=0}^{(\nu_1+\nu_2-1)/2} \int_0^1 a_i x^i dx & \text{if} \quad nS_1 = S_1 + S_2 \\ \sum_{i=1}^{\nu_1+\nu_2} \int_{nS_1}^{S_1+S_2} b_i/x^{i/2} dx & \text{if} \quad nS_1 < S_1 + S_2 \\ \sum_{i=1}^{\nu_1+\nu_2} \int_{S_1+S_2}^{nS_1} b_i/x^{i/2} dx & \text{if} \quad nS_1 > S_1 + S_2 \end{cases}$$

where a_i and b_i are known constants. All moments exist and

$$E(z^r/y) = \begin{cases} C \displaystyle\sum_{i=o}^{(\nu_1+\nu_2-1)/2} \int_o^1 a_i x^{i+r} dx & \text{if } nS_1 = S_1 + S_2 \\ C \displaystyle\sum_{i=1}^{\nu_1+\nu_2} \int_{nS_1}^{S_1+S_2} b_i x^{r-i/2} dx & \text{if } nS_1 < S_1 + S_2 \\ C \displaystyle\sum_{i=1}^{\nu_1+\nu_2} \int_{S_1+S_2}^{nS_1} b_i x^{r-i/2} dx & \text{if } nS_1 > S_1 + S_2 \end{cases}$$

Mode of the Distribution $k \geq 3$, $(n-1)S_1 \neq S_2$

Let $a = (n-1)S_1-S_2$, $b = S_1+S_2$, then differentiating $p(z/y)$ with respect to z we have,

$$\frac{\partial p(z/y)}{\partial z} \; \alpha \; \frac{[1+(n-1)z]^{\frac{k(n-1)-2}{2}} [1-z]^{\frac{k-4}{2}}}{(az+b)^{\frac{kn+1}{2}}} Q(z)$$

where

$$Q(z) = k(n-1)^2(az+b)(1-z)-(k-2)(az+b)[1+(n-1)z]$$
$$- (kn-1)a(1-z)[1+(n-1)z]$$

and

$$\frac{\partial p(z)y}{\partial z} = 0$$

gives

(i) $z = 1 => p(z/y) = 0$ if $k \geq 4$

which is the minimum for a density function. Moreover, $p(z/y) \neq 0$, $0 \leq z < 1$.

(ii) $z = -\frac{1}{n-1}$ outside the range,

(iii) roots of $Q(z) = 0$ as possible maxima or minima.

Collecting the coefficients in the quadratic $Q(z)$, we have

<u>Coefficient of z^2</u>

$$= - ak(n-1)^2-a(n-1)(k-1)+a(n-1)(kn-1)$$

$$= a(n-1)$$

$$= (n-1)^2S_1-(n-1)S_2 \ .$$

<u>Coefficient of z</u>

$$k(n-1)^2(a-b)-(k-2)[a+b(n-1)]-a(kn-1)(n-2)$$

$$= an-b(n-1)(kn-2)$$

$$= -(n-1)[n(k-1)-2]S_1-[n+(n-1)(kn-2)]S_2$$

$$< 0 \quad \text{for all permissible } k,n,S_1,S_2 \ .$$

<u>Constant</u>

$$bk(n-1)^2-b(k-2)-a(kn-1)$$

$$= b[kn(n-2)+2]-a(kn-1) > 0 \quad \text{if} \quad a < 0$$

$$= S_2[kn(n-1)+1]-S_1[n(k-1)-1] \ .$$

The discriminant function of the quadratic is

$$D = [an-b(n-1)(kn-2)]^2-4a(n-1)[b(kn^2-2kn+2)-a(kn-1)]$$

$$= a^2[(n-2)^2+4kn(n-1)]+b^2(n-1)^2(kn-2)^2$$

$$-2ab(n-1)[kn(3n-4)-2(n-2)] \ .$$

Expressing a,b in terms of S_1 and S_2 and collecting coefficient of S_1^2, S_2^2 and S_1S_2 in D we have

<u>Coefficient of S_1^2</u>

$$= (n-1)^2[4kn(n-1)+(kn-2)^2-2kn(3n-4)+(n^2-4)]$$

$$= n^2(n-1)^2(k-1)^2 \ .$$

Coefficient S_2^2

$= 4kn(n-1)+(n-2)^2+(n-1)^2(kn-2)^2+2kn(n-1)(3n-4)-4(n-1)(n-2)$

$= n^2[k(n-1)+1]^2$.

Coefficient of $2S_1S_2$

$= -(n-1)[4kn(n-1)+(n-2)^2]+(n-1)^2(kn-2)^2-(n-1)(n-2)[kn(3n-4)-2(n-2)]$

$= n^2(n-1)[k^2(n-1)-k(3n-2)+1]$

Finally,

$$D = n^2(n-1)^2(k-1)^2S_1^2+n^2[k(n-1)+1]^2S_2^2$$
$$+2S_1S_2n^2(n-1)[k^2(n-1)-k(3n-2)+1]$$
$$= \{n(n-1)(k-1)S_1-n[k(n-1)+1]S_2\}^2+4n^2k(n-1)^2(k-2)S_1S_2 \quad .$$

Hence, $D > 0$ and real distinct roots exist for all permissible values of n,k,S_1 and S_2.

If $a < 0$ i.e. $(n-1)S_1 < S_2$, the constant in $Q(z)$ is positive and the coefficient of z in $Q(z)$ being always negative, $Q(z)$ has the form Az^2-Bz+C where $A,B,C > 0$. By the rule of sign, $Q(z)$ has one positive and one negative real root. If $a > 0$ i.e. $(n-1)S_1 > S_2$, nothing can be said about the constant in $Q(z)$. If the constant is positive, $Q(z)$ has none or two positive real roots and no negative real root. The consideration of the discriminant function has proved the existence of real roots. Hence, $Q(z)$ has two positive roots. If the constant is negative, $Q(z)$ has the form Az^2-Bz-C, $A,B,C > 0$ and has one positive and one negative real root. Thus, we have either one positive root or two positive roots with corresponding one or two possible points of maxima or minima of $p(z/y)$ in the range $0 < z < 1$. The case of the constant in $Q(z)$ being zero is considered separately as a special case.

<u>Case (i)</u> Only one positive real root, a) $S_2/S_1 > n-1$.

When $S_2/S_1 > n-1$, we have seen that the $S_2[kn(n-1)+1]-S_1[n(k-1)-1]$ which is the constant of quadratic $Q(z)$, is positive. Now

$$Q(0) = S_2[kn(n-1)+1]-S_1[n(k-1)-1] > 0$$

$$\begin{aligned} Q(1) &= S_1[(n-1)^2-n(n-1)(k-1)+2(n-1)-n(k-1)+1] \\ &\quad + S_2[kn(n-1)+1-(n-1)-n-(n-1)(kn-2)] \\ &= -n^2(k-2)S_1 < 0 \quad \text{for } k > 2 . \end{aligned}$$

As $Q(z)$ is a continuous and differentiable function of z which changes sign between 0 and 1, $Q(z)$ must vanish for some z_o, $0 < z_o < 1$ and the positive real root lies between 0 and 1 . It is not difficult to show that z_o is the maxima of $p(z/y)$. Thus, posterior distribution of z has a mode in (0,1) and the Bayesian estimate $(\frac{\hat{\sigma}^2_\alpha}{\sigma^2+\sigma^2_\alpha})$ lies between 0 and 1 .

$$\text{b)} \quad \frac{S_2}{S_1} < \frac{n(k-1)-1}{kn(n-1)+1} , \ \frac{S_2}{S_1} < n-1 .$$

$Q(z)$ has the form Az^2-Bz-C, $A,B,C > 0$. We will show that the positive root cannot lie in the interval (0,1) and therefore $p(z/y)$ has no mode and the Bayesian estimate of

$$(\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha})$$

is zero.

$$\frac{\partial p(z/y)}{\partial z}$$

can be written as

$$\frac{\partial p(z/y)}{\partial z} \propto A\psi(z)(z-m_2)(z+m_1) \qquad A > 0$$

where $-m_1$, m_2 are roots of $Q(z)$, satisfying $0 < m_1 < m_2$.

$$\psi(z) = \frac{[1+(n-1)z]^{\frac{k(n-1)-2}{2}} [1-z]^{\frac{k-4}{2}}}{[\{(n-1)S_1-S_2\}z+(S_1+S_2)]^{\frac{kn+1}{2}}}$$

and $\psi(z) > 0$ for $0 \leq z < 1$.

Differentiating $p(z/y)$ twice with respect to z , we have

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_2} A\psi(m_2) \cdot (m_1+m_2) \cdot > 0$$

if $0 < m_2 < 1$.

Suppose that $0 < m_2 < 1$ holds, then m_2 is a minima of $p(z/y)$. We observe that $p(z/y)$ is continuous in the interval $< 0,1 >$ and no where zero except $z = 1$ which is relative minima for $p(z/y)$. As a continuous function can not have two consecutive relative minima, m_2 can not lie in (0,1). At most $m_2 = 1$.

Case (ii) Two positive real roots. $\frac{n(k-1)-1}{kn(n-1)+1} < \frac{S_2}{S_1} < n-1$.

A pertinent question that can be asked in this case is whether the two positive real roots can be in the interval (0,1). Suppose this is possible. Let m_1 , m_2 be roots of $Q(z)$ where $0 < m_1 < m_2 \leq 1$.

$$\frac{\partial p(z/y)}{\partial z} \alpha A\psi(z) \cdot (z-m_1)(z-m_2)$$

where A and $\psi(z)$ are as defined in (i). Differentiating $p(z/y)$ twice with respect to z , we have

$$\left.\frac{\partial p(z/y)}{\partial z^2}\right|_{m_1} \alpha A\psi(m_1)(m_1-m_2) < 0$$

$$\left.\frac{\partial^2 P(z/y)}{\partial z^2}\right|_{m_2} A\psi(m_2)(m_2-m_1) > 0 \quad .$$

Hence m_1 is point of maxima and m_2 is point of minima. We again face the problem of two consecutive minima and therefore m_2 can not be in the interval (0,1). At most $m_2 = 1$. It is possible that both m_1, m_2 may lie outside the range. The posterior distribution of z can have no mode at all and the Bayesian estimate of $\left(\frac{\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2}\right)$ is zero.

Special Case I $\quad (n-1)S_1 = S_2$, $k \geq 4$.

We have

$$p(z/y) \alpha (1-z)^{\frac{k-2}{2}} [1+(n-1)z]^{\frac{k(n-1)}{2}}$$

$$\frac{\partial p(z/y)}{\partial z} \alpha (1-z)^{\frac{k-4}{2}} [1+(n-1)z]^{\frac{k(n-1)-2}{2}} [(kn^2-2kn+2) - (n-1)(kn-2)z]$$

$$\frac{\partial p(z/y)}{\partial z} = 0$$

gives (i) $z_o = 1$, minimum in the range if $k \geq 4$, (ii) $z_o = -\frac{1}{n-1}$ outside the range and (iii)

$$z_o = \frac{kn(n-2)+2}{(n-1)(kn-2)} > 0 \quad ,$$

$$\frac{kn(n-2)+2}{(n-1)(kn-2)} < 1 \quad \text{if} \quad n(k-2) > 0$$

which is satisfied. It is not difficult to show that

$$z_o = (kn^2-2kn+2)/(n-1)(kn-2)$$

is point of maxima of $p(z/y)$. Thus, $p(z/y)$ has a mode in (0,1) and

$$0 < \left(\frac{\hat{\sigma}_\alpha^2}{\sigma^2+\sigma_\alpha^2}\right) < 1 \quad .$$

Special Case II $S_2[kn(n-1)+1] = S_1[n(k-1)-1]$, $k \geq 3$ has the form $\psi(z)\ Az(z-B)$ where $A,B > 0$ and $\psi(z)$ is as defined earlier.

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{z=o} \alpha\text{-}AB\psi(0) < 0$$

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{z=B} \alpha AB\psi(B) > 0,\ 0 < B \leq 1 \quad .$$

Thus, $z = o$ is maxima and $z = B$ is minima. We face the same problem of two consecutive minima and the same remarks apply to $z_o = B$ which can at most be equal to 1.

We summarize the the results as follows:

(i) $\frac{S_2}{S_1} \geq n-1$, $p(z/y)$ has a mode in the interval $(0,1)$.

(ii) $\frac{S_2}{S_1} \leq \frac{n(k-1)-1}{kn(n-1)+1}$, $p(z/y)$ has no mode in the interval $(0,1)$ and

$$\left(\frac{\hat{\sigma}^2_\alpha}{\sigma^2+\sigma^2_\alpha}\right) = 0 \quad .$$

(iii) $\frac{n(k-1)-1}{kn(n-1)+1} < \frac{S_2}{S_1} < n-1$. We can make no definite conclusion and $p(z/y)$ may or may not have a mode in $(0,1)$, depending on n,k,S_1 and S_2 .

(iv) In no case $p(z/y)$ is bi-modal.

E. Posterior Distribution of $\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha}$ with a Reasonable Prior

We have observed earlier that one should base his prior on the history of the case and past experience, however vague it may be. To illustrate this point we consider the following prior, a reasonable one, and inference about $\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha}$ will, then, be based on juidicious utilization of previous knowledge.

$$f(\sigma^2+\sigma^2_\alpha)g(\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha})d(\sigma^2+\sigma^2_\alpha)d(\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha})$$

where f in an inverted scalar χ^2 with ν degrees of freedom and ℓ is a scalar multiplier.

g is a Beta variable with parameters a,b.

The values of ν, ℓ, a and b will depend on the past history of the subject-matter and it is presumed that integral values are taken.

The prior distribution is

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)\Gamma(\frac{\nu}{2})}(\frac{\ell}{2})^{\nu/2}(\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha})^{a-1}(\frac{\sigma^2}{\sigma^2+\sigma^2_\alpha})^{b-1}\frac{1}{(\sigma^2+\sigma^2_\alpha)^{\nu/2+1}} X$$

$$\text{Exp}\{-\tfrac{1}{2}[\frac{\ell}{\sigma^2+\sigma^2_\alpha}]\} .$$

The joint posterior density is

$$p(\sigma^2,\sigma^2_\alpha/y)\alpha\frac{(\sigma^2_\alpha)^{a-1}(\sigma^2)^{b-\nu_1/2-1}}{(\sigma^2+\sigma^2_\alpha)^{a+b+\nu/2-1}(\sigma^2+n\sigma^2_\alpha)^{\nu_2/2}} X$$

$$\text{Exp}\{-\tfrac{1}{2}[\frac{S_1}{\sigma^2}+\frac{S_2}{\sigma^2+n\sigma^2_\alpha}+\frac{\ell}{\sigma^2+\sigma^2_\alpha}]\}d(\sigma^2+\sigma^2_\alpha)d(\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha})$$

writing $\sigma^2+\sigma^2_\alpha = w$, $\frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha} = z$, we have

$$p(w,z/y)\alpha(\frac{1}{w})^{(\nu_1+\nu_2+\nu)/2+1}\frac{z^{a-1}(1-z)^{b-\nu_1/2-1}}{[1+(n-1)z]^{\nu_2/2}} X$$

$$\operatorname{Exp}\{-\frac{1}{2w}[\frac{S_1}{1-z} + \frac{S_2}{1+(n-1)z} + \ell]\}dwdz \qquad w \geq 0,\ 0 \leq z \leq 1 .$$

Integrating w out we have

$$p(z/y)\alpha \frac{z^{a-1}(1-z)^{b-\nu_1/2-1}}{[1+(n-1)z]^{\nu_2/2}} [\frac{S_1}{1-z} + \frac{S_2}{1+(n-1)z} + \ell]^{-(\nu_1+\nu_2+\nu)/2} .$$

Now

$$\frac{S_1}{1-z} + \frac{S_2}{1+(n-1)z} + \ell = \frac{(S_1+S_2+\ell)+z[(n-1)S_1-S_2+(n-2)\ell]-\ell(n-1)z^2}{(1-z)[1+(n-1)z]}$$

$$= \frac{Q(z)}{(1-z)[1+(n-1)z]} \text{ , say.}$$

Then

$$p(z/y)\alpha \frac{z^{a-1}[(1-z)]^{b+(\nu_2+\nu)/2-1}[1+(n-1)z]^{(\nu_1+\nu)/2}}{[Q(z)]^{(\nu_1+\nu_2-\nu)/2}} \qquad 0 \leq z < 1$$

where

$$Q(z) = (S_1+S_2+\ell)+z[(n-1)S_1-S_2+(n-2)\ell]-\ell(n-1)z^2 .$$

As $S_1+S_2+\ell > 0$, $\ell(n-1) > 0$, by the rule of sign, $Q(z)$ has two real roots with opposite signs and can be written as $A(z+B)(C-z)$, $A,B,C > 0$. We will show that for

$$-\frac{1}{n-1} \leq z \leq 1$$

$Q(z) \neq 0$ and therefore $C > 1$ and $B > \frac{1}{n-1}$

$$Q(0) = S_1+S_2+\ell > 0$$

$$Q(1) = (S_1+S_2+\ell)+[(n-1)S_1-S_2+(n-2)\ell]-(n-1)\ell$$

$$= nS_1 > 0$$

$$Q(-\frac{1}{n-1}) = \frac{1}{n-1}[(n-1)(S_1+S_2+\ell)-(n-1)S_1+S_2-(n-1)\ell]$$
$$= \frac{nS_2}{n-1} > 0 \quad .$$

$Q(z)$ vanishes once only somewhere on the positive axis and once only somewhere on the negative axis and the maximum value of $Q(z)$ is positive. The positive values at two points imply that $Q(z)$ is positive between those points. We can therefore conclude that $Q(z)$ does not vanish between $-\frac{1}{n-1}$ and 1. Thus, $p(z/y)$ as a function of z is well defined over the range $-\frac{1}{n-1} \leq z \leq 1$ and can be so extended. This fact will be exploited while considering mode of $p(z/y)$.

The integration of $p(z/y)$ for the normalizing constant, in general, is very complicated. Numerical methods are available to calculate the integral. However, under the following assumptions, the integral can be easily calculated.

(i) ν is even,

(ii) $\nu_1 = k(n-1)$ and $\nu_2 = k-1$ are even, i.e. the number of classes as well as the number of observations per class are odd. Then

$$p(z/y) \alpha \frac{P^{a+b+\nu+(\nu_1+\nu_2)/2-2}(z)}{[(z+B)(C-z)]^{(\nu_1+\nu_2+\nu)/2}}$$

where P^r denotes a polynomial of degree r.

Letting $z + B = x$ (we can put $C-z = x$ also)

$$\int_0^1 \frac{P^{a+b+\nu+(\nu_1+\nu_2)/2-2}(z)}{[(z+B)(C-z)]^{(\nu+\nu_1+\nu_2)/2}} dz \qquad B > \frac{1}{n-1}, \; C > 1$$

$$= \int_{B}^{1+B} \frac{P^{a+b+\nu+(\nu_1+\nu_2)/2-2}(x)}{[x(x-A)]^{(\nu+\nu_1+\nu_2)/2}}\, dx \quad \text{where} \quad A = B + C \ .$$

It involves the evaluation of the integrals of the following forms.

$$I = \int_{B}^{1+B} \frac{x^m}{(x-A)^n}\, dx = \int_{-C}^{1-C} \frac{(A+y)^m}{y^n}\, dy$$

$$II = \int_{B}^{1+B} \frac{dx}{(x-A)^n}\, dx$$

$$III = \int_{B}^{1+B} \frac{dx}{x^m(x-A)^n}\, dx$$

where m and n are known positive integers.

The evaluation of above integrals is not difficult. I and II can be evaluated directly. III can be solved through integration by parts.

1. Mode of the posterior distribution

Let $p = \ell(n-1)$, $q = (n-1)S_1-S_2+(n-2)\ell$, $r = S_1+S_2+\ell$,

$$p(z/y)\alpha\ \frac{z^{a-1}(1-z)^{(2b+\nu+\nu_2-2)/2}[1+(n-1)z]^{(\nu_1+\nu)/2}}{(r+qz-pz^2)^{(\nu_1+\nu_2+\nu)/2}} \ .$$

Differentiating $p(z/y)$ with respect to z and equating to zero, we have

$$\frac{z^{a-2}(1-z)^{(2b+\nu+\nu_2)/2-2}[1+(n-1)z]^{(\nu_1+\nu)/2-1}}{(r+qz-pz^2)^{(\nu_1+\nu_2+\nu)/2+1}} \times$$

$$\begin{bmatrix} 2(a-1)(1-z)(r+qz-pz^2)\{1+(n-1)z\}+(n-1)(\nu_1+\nu)z(1-z)(r+qz-pz^2)] \\ -(2b+\nu+\nu_2-2)(r+qz-pz^2)\{z+(n-1)z^2\}+(\nu_2+\nu_1+\nu)(2pz-q)(1-z)\{z+(n-1)z^2\} \end{bmatrix} = 0$$

which gives $z = 0$ if $a > 2$, $z = 1$ and the roots of the quartic $QR(z)$. as possible points of maxima or minima. But $z = 0$ and $z = 1$ give $p(z) = 0$ which is the minima for $p(z/y)$ in the range $0 \leq z \leq 1$. We are not interested in $z = -(\frac{1}{n-1})$, being outside the range. For the mode, we consider the roots of the quartic $QR(z)$.

For convenience, we collect the coefficients of the quartic in two stages.

1st stage

$$(r+qz-pz^2)[2(a-1)(1-z)\{1+(n-1)z\}+(n-1)(\nu_1+\nu\ )z(1-z)$$
$$+(2b+\nu+\nu_2-2)z\{1+(n-1)2\}$$

$$=(r+qz-pz^2)[2(a-1)+\{(n-2)(2a+\nu+kn-2)-(2b-3)\}z$$
$$-(n-1)(2a+2b+2\nu+kn-5)z^2] .$$

Coefficients are

z^4 : $p(n-1)(2a+2b+2\nu+kn-5)$

z^3 : $-p[(n-2)(2a+\nu+kn-2)-(2b-3)]-q(n-1)(2a+2b+2\nu+kn-5)$

z^2 : $q[(n-2)(2a+\nu+kn-2)-(2b-3)]-r(n-1)(2a+2b+2\nu+kn-5)-2p(a-1)$

z : $2(a-1)q+[(n-2)(2a+\nu+kn-2)-(2b-3)]r$

and

Const: $2(a-1)r$.

2nd stage

$$(\nu_1+\nu_2+\nu)(2pz-q)z(1-z)[1+(n-1)z]$$

coefficients are

z^4 : $-2p(n-1)(kn+\nu-1)$

z^3 : $[2p(n-2)+(n-1)q](kn+\nu-1)$

$$z^2 \; : \; [2p-q(n-2)](kn+\nu-1)$$

$$z \; : \; -q(kn+\nu-1)$$

and

$$\text{Const:} \;\; \text{zero} \; .$$

Finally the coefficients of the quartic are

$$z^4 \; : \; p(n-1)(2a+2b-kn-3)$$

$$z^3 \; : \; p[(n-2)(kn+\nu-2a)+(2b-3)]-q(n-1)(2a+2b+\nu-4)$$

$$z^2 \; : \; 2p(kn+\nu-a)+q[(n-2)(2a-1)-(2b-3)]-r(n-1)(2a+2b+2\nu+kn-5)$$

$$z \; : \; r[(n-2)(2a+\nu+kn-2)-(2b-3)]-q[kn+\nu-2a+1]$$

and

$$\text{Const:} \; 2(a-1)r > 0$$

where

$$p = \ell(n-1) > 0,$$

$$q = (n-1)S_1-S_2+(n-2)\ell,$$

$$r = S_1+S_2+\ell > 0 \; .$$

We can extend $p(z/y)$ as a continuous function of z over the range $-\frac{1}{n-1} \leq z \leq 1$. $p(z/y)$ has the value zero at $z = 0$ and $z = 1$ and positive within the open interval $(0,1)$. Hence, $p(z/y)$ has at least one mode in that interval and the quartic has at least one positive real root between 0 and 1. Considering $p(z/y)$ as a function in the closed interval $< -\frac{1}{n-1}, 0 >$, we observe that $p(z/y)$ is zero at $z = 0$ and $z = -\frac{1}{n-1}$ and has the same sign within the interval (depending on values of parameters). We, therefore, conclude that the quartic has at least one negative real root between $-\frac{1}{n-1}$ and 0. Thus the quartic has at least two real roots with opposite signs.

Suppose that $p(z/y)$ is not uni-modal. As $p(z/y)$ is zero at $z = 0$ and $z = 1$ and positive within the interval $(0,1)$, $p(z/y)$ must have successively maxima, minima and maxima and the quartic has all real roots. Let $0 < m_1 < m_2 < m_3 < 1$ be the positive real roots and $-\frac{1}{n-1} < m_4 < 0$ be the negative real root, then the quartic has the form

$$A(z-m_1)(z-m_2)(z-m_3)(z+m_4)$$

where

$$A = \ell(n-1)^2(2a+2b-kn-3)$$

and

$$A(-m_1)(-m_2)(-m_3)(m_4) = 2(a-1)(S_1+S_2+\ell) > 0 \quad .$$

It is therefore necessary that $A < 0$ i.e. $2(a+b) < (kn+3)$. The condition is not sufficient because with $A < 0$, m_1 and m_2 may be outside the range and also m_1 and m_2 may both be negative.

$\frac{\partial p(z/y)}{\partial z}$ can be written as

$$\frac{\partial p(z/y)}{\partial z} \propto \psi(z)A(z-m_1)(z-m_2)(z-m_3)(z+m_4) \qquad A < 0$$

where

$$\psi(z) = \frac{z^{a-2}(1-z)^{b+\frac{\nu_2+\nu}{2}-2}[1+(n-1)z]^{(\nu+\nu_1)/2-1}}{(r+qz-pz^2)^{(\nu_1+\nu_2+\nu)/2-1}}$$

$$\psi(z) > 0 \quad \text{for} \quad 0 < z < 1$$

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_1} = A\psi(m_1)(m_1-m_2)(m_1-m_3)(m_1+m_4) \quad < 0$$

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_2} = A\psi(m_2)(m_2-m_1)(m_2-m_3)(m_3+m_4) \quad > 0$$

$$\left.\frac{\partial^2 p(z/y)}{\partial z^2}\right|_{m_3} = A\psi(m_3)(m_3-m_1)(m_3-m_2)(m_3+m_4) \quad < 0$$

Thus, m_1, m_2, m_3 are points of maxima, minima and maxima respectively and with $kn+3 > 2(a+b)$ it is possible to have $p(z/y)$ bi-modal.

If $2(a+b) = kn+3$, the quartic is reduced to cubic and $p(z/y)$ cannot have more than one mode. If $2(a+b) > (kn+3)$ then other two real roots must exist and have opposite signs. We have, therefore, two positive real roots, one in the interval (0,1). The other positive real root cannot be in (0,1) as $p(z/y)$ is zero at $z = 0$ and $z = 1$ and no where zero. At most it can be equal to 1. Thus, $p(z/y)$ is uni-modal. This is a sufficient condition, but not necessary. We have seen that with $2(a+b) < (kn+3)$ and m_1, m_2 outside the range, $p(z/y)$ can be uni-modal.

Thus, $(kn+3) < 2(a+b)$ is a sufficient, but not necessary, condition for $p(z/y)$ to be uni-modal and $(kn+3) > 2(a+b)$ is a necessary, but not sufficient, condition for $p(z/y)$ to be bi-modal.

Example 1 $(kn+3) < 2(a+b)$

Suppose that we have a prior knowledge that $z = \sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$ has a distribution with mean 0.4 and variance .01 and the gross variability $(\sigma^2+\sigma_\alpha^2)$ has a distribution with mean 80 and variance 400. Equating first two moments of the beta distribution with prior knowledge of mean and variance, we have the prior distribution of z as $\beta(9,14)$. Similarly, $(\sigma^2+\sigma_\alpha^2)$ is distributed as $2720/\chi^2_{34}$. Let the sample observations be

$$n = 3,\ k = 7,\ S_1 = S_2 = 2720$$

which satisfy $(kn+3) < 2(a+b)$.

The usual Analysis of Variance table is shown below.

Table 2. Analysis of Variance

Due to	d-f.	S.S	M.SS.	E.M.SS.	F	
Between	6	2720	453.33	$\sigma^2+3\sigma_\alpha^2$	2.33	not sig.
Within	14	2720	194.29	σ^2		

The parameters are

$$a = 9,\ b = 14,\ \nu_1 = 14,\ \nu_2 = 6,\ \nu = 34,$$

$$S_1 = S_2 = \ell = 2720 \quad .$$

The posterior distribution of z is

$$p(z/y) \propto \frac{z^8(1-z)^{33}(1+2z)^{24}}{[(1+z)(3-z)]^{27}} \qquad 0 \le z \le 1 .$$

and

$$\frac{\partial p(z/y)}{\partial z} \propto \frac{z^7(1-z)^{32}(1+2z)^{23}}{[(1+z)(3-z)]^{28}} (22z^4-121z^3-352z^2+31z+24)$$

$$\propto \psi(z)P^4(z)$$

By the rule of sign the quartic $P^4(z)$ has two or none positive real roots and two or none negative real roots. It is obvious that $p(z/y)$ as a function of z can be extended over the interval $-\frac{1}{2} \le z \le 1$.

We consider $\overset{4}{P}(z)$ in the extended region

$$\overset{4}{P}(1) = -396$$
$$\overset{4}{P}(0) = 24$$
$$\overset{4}{P}(-\tfrac{1}{2}) = -63 \ .$$

$\overset{4}{P}(z)$ changes signs between $(0,1)$ and $(-\frac{1}{2},0)$ and therefore, has at least one positive and one negative real root. In view of the rule of sign, $\overset{4}{P}(z)$ has all real roots -- two positive and two negative. We are not interested in negative real roots for the mode of $p(z/y)$. Of the two positive real roots, one lies in the interval $(0,1)$. The other positive real root cannot lie in the interval $(0,1)$ for reasons explained earlier. Factorizing $\overset{4}{P}(z)$ we have

$$\overset{4}{P}(z) = 22(z-0.293)(z-2.047)(z+0.229)(z+7.611) \ .$$

Hence, the mode of z is 0.293 and Bayesian estimate of $\sigma^2_\alpha/(\sigma^2+\sigma^2_\alpha)$ is 0.293.

As the mode of $\beta(a,b)$ is at $\frac{a-1}{a+b-2}$ the prior estimate of z is 0.381.

Example 2 $kn+3 > 2(a+b)$

Suppose that prior knowledge about $\sigma^2_\alpha/(\sigma^2+\sigma^2_\alpha)$ and $(\sigma^2+\sigma^2_\alpha)$ discussed in Example 2 holds good and our sample observations are

$n = 5$, $k = 9$ so that $kn + 3 > 2(a+b)$, $\ell = S_1 = S_2 = 2720$, $\nu = 34$.

The usual analysis of variance of data is shown in Table 3.

The posterior distribution of z is

$$p(z/y) \ \alpha \ \frac{z^8(1-z)^{34}(1+4z)^{35}}{(3+6z-4z^2)^{39}} \qquad 0 \le z \le 1$$

$$\alpha \frac{z^8(1-z)^{34}(1+4z)^{35}}{[(1.8956-z)(z+0.3956)]^{39}} \qquad 0 \le z \le 1 .$$

Table 3. Analysis of Variance of the data

Due to	d-f.	S.S	M.SS.	E.M.SS.	F
Between	8	2720	340.00	$\sigma^2+5\sigma_\alpha^2$	4.50 highly sig.
Within	36	2720	75.56	σ^2	

$$\hat{\sigma}^2 = 75.56,\ \hat{\sigma}_\alpha^2 = 52.89,\ \left(\frac{\hat{\sigma}_\alpha^2}{\sigma^2+\sigma_\alpha^2}\right) = 0.412 .$$

It may be noticed that $p(z/y)$ as a function of z can be extended over the interval $-0.25 \le z < 1.8956$ and

$$\frac{\partial p(z/y)}{\partial z} \alpha \frac{z^7(1-z)^{33}(1+4z)^{34}}{[(1.8956-z)(z+0.3956)]^{40}} \cdot \overset{4}{P}(z)$$

where $\overset{4}{P}(z) = 24 + 204z-566z^2-496z^3-16z^4$.

By the rule of sign $\overset{4}{P}(z)$ has only one positive real root and three negative real roots. As $\overset{4}{P}(0) = 12$ and $\overset{4}{p}(1) = -425$, the positive real root lies between $(0,1)$ and $p(z/y)$ is uni-modal. This example illustrates that $(kn+3) > 2(a+b)$ is not a sufficient condition for $p(z/y)$ to be bi-modal. It is only a necessary condition.

Factorizing $\overset{4}{P}(z)$, we have

$$\overset{4}{P}(z) = 8(0.362-z)(z+0.094)(z+1.468)(z+29.779) .$$

The posterior estimate is $\left(\frac{\hat{\sigma}_\alpha^2}{\sigma^2+\sigma_\alpha^2}\right) = 0.362$.

A comparison of posterior estimate of z and sample theory estimate of z in the above examples is interesting.

Estimator	Example I	Example II
Posterior (z)	0.293	0.362
Sample Theory (z)	0.308	0.412

It appears that the difference between two estimates tend to increase as the F-value increases. In the first example, observed F is not significant and the difference is small. In the second example, observed F is highly significant and the difference is large. Is there any theoretical reason for this? Further investigations are required.

F. Examples of Posterior Modes and Means

For the purpose illustrating the results by numerical examples and graphs, we have constructed an artificial set of data. We have considered the following samples sizes.

(i) $k = 4$, $n = 5$

(ii) $k = 10$, $n = 5$

(iii) $k = 20$, $n = 5$

where

k = number of groups,

n = number of observations in each group.

For each sample size, we have chosen five sub-sets of 'within' sum of squares, denoted by S_1 , and 'between' sum of squares, denoted by S_2 , in such a way that error mean square is always 100 and the ratios $\frac{\text{M.S. 'between'}}{\text{M.S. 'within'}}$

equal (i) 0.25, (ii) 0.50, (iii) 0.75, (iv) 0.95 and (v) 0.99 cumulative F-value. Thus we have fifteen examples covering a wide range of F-ratios and small to large sample sizes. It is hoped that the examples fairly represent the types of cases we meet frequently. The artificial sets of data, which are used throughout the present study, are given in Appendix A.

It may be recalled that we have studied the posterior distributions of (i) $z = (\sigma^2+\sigma_\alpha^2)/\sigma^2$ and (ii) $z = \sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$ using non-informative priors (i) $\frac{d\sigma^2}{\sigma^2} \cdot \frac{d\tau^2}{\tau^2}$ and (ii) $\frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau}$ for each , where $\tau^2 = \sigma^2 + \sigma_\alpha^2$. Thus, we have four distributions. For each posterior distribution, we have calculated (i) mode, (ii) mean , and (iii) A.O.V. estimate of z, using the artificial set of data containing fifteen numerical examples. The integration for normalizing constants and means was done on 360-65 IBM Computer at Iowa State University, Ames, Iowa. The evaluation of integrals was done by means of trapezoidal rule according to Pomberg's principle. The analysis of variance estimate (A.O.V.E.) was calculated by substituting the estimates of σ^2 and σ_α^2 as obtained from A.O.V. table for σ^2 and σ_α^2 in $(\sigma^2+\sigma_\alpha^2)/\sigma^2$ and $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$. The results are presented in Tables 4, 5, 6, and 7.

The traditional estimate of σ_α^2 is zero when the posterior distribution has no mode in the range over which it is defined. The mode and A.O.V. estimate, therefore, coincide when the cumulative F-value is 0.25 or 0.50 and are equal to the lower end of the range. Hence, the posterior mode and the A.O.V. estimate of $(\sigma^2+\sigma_\alpha^2)/\sigma^2$ is 1.000 for the cumulative F-value 0.25 and 0.50, whatever may be the prior and the sample size. Similarly for $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$ both the posterior mode and the A.O.V. estimate are zero.

It is interesting to note that the posterior mode with the prior $\frac{d\sigma^2}{\sigma^2} \cdot \frac{d\tau^2}{\tau^2}$ is generally higher than the posterior mode with the prior $\frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau}$, both for $(\sigma^2+\sigma^2_\alpha)/\sigma^2$ and $\sigma^2_\alpha/(\sigma^2+\sigma^2_\alpha)$ and for all sample sizes. This is due to the difference in weights attached to points in the parameter space by the priors. The prior $\frac{d\sigma^2}{\sigma^2} \cdot \frac{d\tau^2}{\tau^2}$ gives more weight to points than the prior $\frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau}$. This causes the posterior mode to shift more towards the right side for the first prior than the second prior.

A comparative study of means and modes of the posterior distributions shows that for small sample size $(k = 4,\ n = 5)$, the difference between mean and the corresponding mode is large. The higher cumulative F-values do not appear to have any striking effect in reducing the difference. For moderately large sample size and moderate cumulative F-value, the difference is not large and tends to decrease with large sample size and high cumulative F-values. The A.O.V. estimates are sufficiently close to mean and mode under these conditions.

Table 4. Posterior distribution of $(\sigma^2+\sigma_\alpha^2)/\sigma^2$, using the prior $\frac{d\sigma^2}{\sigma^2}\ \frac{\sigma\tau^2}{\tau^2}$ where $\tau^2 = \sigma^2 + \sigma_\alpha^2$

	Cumulative F-value				
	0.25	0.50	0.75	0.95	0.99
			(i) $k = 4$, $n = 5$		
Normalizing Constant	$\frac{10^{37}}{581,983}$	$\frac{10^{37}}{434,139}$	$\frac{10^{37}}{292,247}$	$\frac{10^{37}}{145,017}$	$\frac{10^{38}}{841,345}$
Mode	1.000	1.000	1.053	1.292	1.552
Mean	1.846	2.259	2.359	3.829	4.027
A.O.V.E.[1]	1.000	1.000	1.103	1.448	1.859
			(ii) $k = 10$, $n = 5$		
Normalizing Constant	$\frac{10^{96}}{247,311}$	$\frac{10^{97}}{981,385}$	$\frac{10^{97}}{334,185}$	$\frac{10^{98}}{645,287}$	$\frac{10^{98}}{196,345}$
Mode	1.000	1.000	1.051	1.189	1.319
Mean	1.152	1.193	1.275	1.466	1.672
A.O.V.E.[1]	1.000	1.000	1.068	1.224	1.376
			(iii) $k = 20$, $n = 5$		
Normalizing Constant	$\frac{10^{202}}{131,686}$	$\frac{10^{203}}{248,185}$	$\frac{10^{204}}{405,267}$	$\frac{10^{205}}{265,053}$	$\frac{10^{206}}{375,746}$
Mode	1.000	1.000	1.041	1.131	1.212
Mean	1.071	1.096	1.135	1.230	1.327
A.O.V.E.[1]	1.000	1.000	1.048	1.144	1.230

[1]A.O.V.E = Analysis of Variance estimate.

Table 5. Posterior distribution of $(\sigma^2+\sigma_\alpha^2)/\sigma^2$, using the prior $\frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau}$ where $\tau^2 = \sigma^2 + \sigma_\alpha^2$

	Cumulative F-value				
	0.25	0.50	0.75	0.95	0.99
			(i) k = 4 , n = 5		
Normalizing Constant	$\frac{10^{37}}{482,875}$	$\frac{10^{37}}{348,472}$	$\frac{10^{37}}{222,691}$	$\frac{10^{38}}{984,350}$	$\frac{10^{38}}{512,257}$
Mode	1.000	1.000	1.036	1.245	1.466
Mean	1.560	1.686	1.904	2.478	3.155
A.O.V.E.[1]	1.000	1.000	1.103	1.448	1.859
			(ii) k = 10 , n = 5		
Normalizing Constant	$\frac{10^{96}}{234,522}$	$\frac{10^{97}}{907,510}$	$\frac{10^{97}}{300,452}$	$\frac{10^{98}}{545,852}$	$\frac{10^{98}}{156,786}$
Mode	1.000	1.000	1.044	1.174	1.296
Mean	1.133	1.177	1.249	1.419	1.600
A.O.V.E.[1]	1.000	1.000	1.068	1.224	1.376
			(iii) k = 20 , n = 5		
Normalizing Constant	$\frac{10^{202}}{127,460}$	$\frac{10^{203}}{237,628}$	$\frac{10^{204}}{381,716}$	$\frac{10^{205}}{240,504}$	$\frac{10^{206}}{328,906}$
Mode	1.000	1.000	1.037	1.125	1.203
Mean	1.068	1.092	1.130	1.219	1.312
A.O.V.E.[1]	1.000	1.000	1.048	1.144	1.230

[1]A.O.V.E. = Analysis of Variance Estimate.

Table 6. Posterior distribution of $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$, using the prior $\frac{d\sigma^2}{\sigma^2} \cdot \frac{d\tau^2}{\tau^2}$ where $\tau^2 = \sigma^2 + \sigma_\alpha^2$

	Cumulative F-value				
	0.25	0.50	0.75	0.95	0.99
			(i) k = 4 , n = 5		
Normalizing Constant	$\frac{10^{37}}{581,983}$	$\frac{10^{37}}{434,139}$	$\frac{10^{37}}{292,247}$	$\frac{10^{37}}{145,017}$	$\frac{10^{38}}{841,345}$
Mode	0.000	0.000	0.171	0.486	0.676
Mean	0.285	0.326	0.386	0.503	0.594
A.O.V.E.[1]	0.000	0.000	0.093	0.309	0.462
			(ii) k = 10 , n = 5		
Normalizing Constant	$\frac{10^{96}}{247,311}$	$\frac{10^{97}}{981,385}$	$\frac{10^{97}}{334,185}$	$\frac{10^{98}}{645,287}$	$\frac{10^{98}}{196,345}$
Mode	0.000	0.000	0.081	0.213	0.312
Mean	0.111	0.140	0.185	0.275	0.351
A.O.V.E.[1]	0.000	0.000	0.064	0.183	0.273
			(iii) k = 20 , n = 5		
Normalizing Constant	$\frac{10^{202}}{131,686}$	$\frac{10^{203}}{248,185}$	$\frac{10^{204}}{405,267}$	$\frac{10^{205}}{265,052}$	$\frac{10^{206}}{375,746}$
Mode	0.000	0.000	0.053	0.137	0.201
Mean	0.062	0.082	0.111	0.173	0.230
A.O.V.E.[1]	0.000	0.000	0.046	0.126	0.187

[1] A.O.V.E. = Analysis of Variance Estimate.

Table 7. Posterior distribution of $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$, using the prior $\frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau}$ where $\tau^2 = \sigma^2 + \sigma_\alpha^2$

	Cumulative F-value				
	0.25	0.50	0.75	0.95	0.99
			(i) k = 4 , n = 5		
Normalizing Constant	$\frac{10^{37}}{482,875}$	$\frac{10^{37}}{348,472}$	$\frac{10^{37}}{222,691}$	$\frac{10^{38}}{984,350}$	$\frac{10^{38}}{512,257}$
Mode	0.000	0.000	0.125	0.376	0.545
Mean	0.240	0.275	0.328	0.436	0.526
A.O.V.E.[1]	0.000	0.000	0.093	0.309	0.462
			(ii) k = 10 , n = 5		
Normalizing Constant	$\frac{10^{96}}{234,522}$	$\frac{10^{97}}{907,510}$	$\frac{10^{97}}{300,452}$	$\frac{10^{98}}{545,852}$	$\frac{10^{98}}{156,786}$
Mode	0.000	0.000	0.072	0.197	0.292
Mean	0.104	0.132	0.173	0.258	0.329
A.O.V.E.[1]	0.000	0.000	0.064	0.183	0.273
			(iii) k = 20 , n = 5		
Normalizing Constant	$\frac{10^{202}}{127,460}$	$\frac{10^{203}}{237,628}$	$\frac{10^{204}}{381,716}$	$\frac{10^{205}}{240,504}$	$\frac{10^{206}}{328,906}$
Mode	0.000	0.000	0.049	0.131	0.194
Mean	0.061	0.080	0.107	0.167	0.222
A.O.V.E.[1]	0.000	0.000	0.046	0.126	0.187

[1]A.O.V.E. = Analysis of Variance Estimate.

IV. LIKELIHOOD INFERENCE

A. General Discussion

The concept of likelihood is due to Fisher. He uses likelihood as a predicate of an hypothesis in the light of data. The term data includes the specification of the problem i.e. the model assumed as well as the observations taken. If the distribution of chances admits a continuous or discrete density function represented by $f(x,\theta)$, then $f(x,\theta)$ as a function of θ is the likelihood of θ. The likelihood is to be distinguished from the probability. The likelihood does not obey the Kolmogoroff's axioms in the sense that the sum of likelihoods taken over mutually exclusive hypotheses or over continuously many possible hypotheses is not 1, while the sum of probabilities over the sample space for a given θ is 1. This is due to the fact that the likelihood does not involve a measurable reference set as the probability does. Another distinguishing feature of the likelihood is the Jacobian of a transformation of the parameter space, which does not enter into the likelihood. If we change from θ to ψ, then we need only to replace θ by $\theta(\psi)$ in the likelihood. The Jacobian of the transformation $|d\theta/d\psi|$ is not incorporated in the likelihood, but the same is incorporated in a probability density function.

The likelihood function can be used to determine the relative plausibility of two competing hypotheses in the light of data or it may be used as a measure of relative belief. If θ_1 and θ_2 are the two hypothesized values of the parameter θ, then the ratio $L(\theta_1)/L(\theta_2)$ measures the relative degree of belief in the hypotheses or the extent to which the evidence (data) favors θ_1 as compared with θ_2 . We can also use likelihoods to give an ordering to θ_1 and θ_2 and in general to various values of θ in the parameter space. A single value of a likelihood says nothing about θ . It may be a very small quantity, suggesting that a rare event has happened, but we do not know if there exists another likelihood value which suggests the happening of a more rare event. The likelihood is, therefore, used as a ratio.

Barnard is one of the most vocal exponents of the use of likelihood in inference problems. He advocates that before the experiment probabilities are relevant and after the experiment likelihoods are relevant. It would follow that where the likelihood principle is applicable, the primary source of inference should be the likelihood function. Barnard et al. (2) consider experimental situations which can be expressed in terms of (i) a sample space, (ii) a parameter space and (iii) a function of two variables, called the kernel. The first variable of the kernel ranges over the sample space and the second variable ranges over the parameter space. An experimental situation can, therefore, be represented by a triplet (S,Ω,f) where S is the sample space, Ω is the parameter space and f is the kernel. If we observe a result $x \in S$ and for a given $\theta \in \Omega$, $f(x,\theta)$ is the kernel, then $f(x,\theta)$ is the likelihood of θ , given x . The likelihood can be used as a primary source of inference only when the triplet (S,Ω,f) specifies all the inferential features of the experimental situation.

Suppose that (S,Ω,f) and (T,Ω,g) are two experimental situations with a common parameter space. Let x be any result from (S,Ω,f) and y be any result from (T,Ω,g). If there exists a positive constant c such that $f(x,\theta) = cg(y,\theta)$ for all $\theta \in \Omega$, then the likelihood functions

$f(x,\theta)$ and $g(y,\theta)$ are said to be equivalent or the same. The likelihood principle states that the inference from x about θ would be the same as the inference from y about θ. In other words, the inference is characterized completely by the likelihood function and is independent of the structure of the experiment otherwise. The following example given by Barnard et al. (2) illustrates the point.

Let the given data be: 3 heads out of 20 throws of a penny. Consider the experimental situations (i) the penny was tossed 20 times and 3 heads were observed and (ii) the intention was to toss the penny till 3 heads are observed and the penny was tossed 20 times. If θ is the chance of a head in a single throw, then the parameter space $(0 \leq \theta \leq 1)$ is the same in the two experiments, but the sample spaces differ. We observe that the likelihood of θ for (i) is $1140\ \theta^3(1-\theta)^{17}$ and for (ii) it is $171\ \theta^3(1-\theta)^{17}$. The likelihoods are equivalent, and therefore, if one accepts the "likelihood principle", the inference about θ should be the same. The stopping rule is irrelevant here as both the rules yield equivalent likelihoods. It may be noted that the maximum likelihood estimate of θ is 0.15 for both the experiments.

The likelihood principle is not compatible with significance and confidence interval procedures. These procedures will yield different results for the two experiments considered above. This is due to the fact that for the application of the significance and confidence interval procedures, we must know the totality of possible results with which the given result is to be compared. No such reference set is involved in the likelihood principle. As the sample spaces are different, these procedures may yield different results.

To sum up, the likelihood principle enables us to make any pairwise comparison of two points in the parameter space. We can select any pair of points (θ_1, θ_2) in the parameter space and compare $L(\theta_1)$ with $L(\theta_2)$ to draw an appropriate inference. The likelihood function is not additive and we can not, in general, compare a set of points in the parameter space with another. We can not, therefore, calculate the likelihood of $(a < \theta < b)$ where a and b are fixed constants belonging to the parameter space. Any statement similar to confidence limit is not, generally, possible. However, exceptions are possible. If the sample space S and the parameter space Ω, have some ordering structure or group structure or special features, then it may be reasonable to make a departure from the likelihood principle and apply a form of argument which is, otherwise, not applicable. The experimenter will have to show that the special features of his experiment justify his special argument. For example, if both the parameter and sample spaces have group structures, then it may be possible to integrate the likelihood function to give something like confidence distribution from which confidence statements may be derived. It may, however, be noted that this implies a departure from the likelihood principle which may be justified, if at all, only under special conditions.

Closely connected with the principle of likelihood is the principle of conditionality which is also due to Fisher, and has been examined in detail by Birnbaum (3). He considers a two-stage experiment E with components $E_h = (S_h, \Omega, f_h)$. In the first stage, an observation h is taken at random according to some distribution G defined over $\{h\}$ which is independent of the parameter. In the second stage, the corresponding experiment E_h is conducted and an outcome x_h is observed. Such an experiment E is called a mixture experiment and its possible outcomes are denoted by $(E,(E_h,x_h))$. The principle of conditionality states that for the purpose of inference about the unknown parameter, $(E,(E_h,x_h))$ and (E_h,x_h) are equivalent. The over-all structure of the experiment E is ignored. A direct implication of the principle of conditionality is the principle of sufficiency which states that if two outcomes of the same experiment determine equivalent likelihoods, then they yield the same inference. The likelihood principle and the conditionality principle are equivalent and they imply the sufficiency principle. For details, please see Kempthorne's comments and Birnbaum's reply in (5).

The basis of the theory of support advocated by Hacking (16) is the likelihood concept. A brief review follows. He defines a statistical hypothesis as an hypothesis about the distribution of outcomes of trials of a specified kind on some chance set-up. The concept of a chance set up is vital in his theory. We use his own words. "A chance setup is a device or part of the world on which might be conducted one or more trials, experiments or observations; each trial must have a unique result, which is a member of a class of possible results." The chance or the long run frequency or the probability is the property of a chance setup. A piece of radium with a recording mechanism or a coin with a coin tossing device may provide a chance setup. A trial may be observing the amount of radium emitted in a specified time interval or tossing the coin thrice and observing the number of heads. Given data, the question is which of the several hypotheses is best supported. His answer is in terms of "simple joint proposition" and "joint position" which need a little explanation.

A simple joint proposition is defined as a conjunction of two propositions, (a) a single distribution of chances of outcomes of trials of a specified kind on a given chance setup, and (b) that a specified outcome occurs on a designated trial of that kind. The likelihood of a simple joint proposition is the chance of outcome (b) if the distribution (a) is true. This is a number assigned to a simple joint proposition specified by (a) and (b). A joint proposition is a conjunction of the two propositions, (c) an hypothesis that the distribution of chances on outcomes of trials of some kind on a chance setup belongs to some specified class of distribution, and (d) that a specified outcome occurs on trials of some (possibly different) kind on the same setup. A simple joint proposition h is said to be included in a joint proposition e if e is logically equivalent (one entailing the other) to a joint proposition e' such that the distribution specified by h is a member of the class of distributions specified e' and the outcome of trials of the kind specified by h is contained in the outcome of trials of the same kind by e'. The truth or otherwise of a joint proposition is not to be questioned. The inference is built up on the joint proposition. A joint proposition may be a statement about some statistical data which includes the model assumed and observations taken. A simple joint proposition may specify a hypothesized value of the parameter.

Let d be a joint proposition which includes simple joint propositions h and i. The simple joint propositions yield their respective likelihoods. The law of likelihood, enunciated by Hacking (16) states that d supports h better than i if the likelihood ratio of h to i exceeds 1. The law of likelihood can be re-stated in another way. The likelihood of a simple joint proposition, given a joint proposition d which includes h, is the absolute likelihood of h . This is a number assigned to h . This concept implies an equivalence condition. If h and i, given d, are equivalent, then the likelihoods of h and i, given d are the same. The second version says that d supports h better than i if the likelihood of h, given d, exceeds the likelihood of i, given d . Evidently, Hacking measures the relative support of two hypotheses, consistent with data, in terms of relative likelihood. Hence any part of likelihood which is independent of parameter θ is irrelevant because it cancels in a ratio.

Hacking's theory of support seems to be in agreement with the principle of likelihood. There is a difference in the method of expression. Suppose that d and e are joint propositions or two experimental situations. Let h and i be hypotheses which are included in d as well as e . In the language of the principle of likelihood we say that h and i hypothesize values of the parameter, θ_1 and θ_2 , in the common parameter space. Suppose that the likelihoods of h and i, given d , are equivalent. If $p(h/d)$ denotes the support of h given by d, then there exists a positive constant c such that $p(h/d) = cp(h/e)$ and $p(i/d) = cp(i/e)$, and hence

relative support of h to i , given d

$$= \frac{p(h/d)}{p(i/d)} = \frac{cp(h/e)}{cp(i/e)} = \frac{p(h/e)}{p(i/e)}$$

= relative support of h to i , given e .

Obviously the inference is the same and is, otherwise, independent of the structure of the experiment.

In the present study, we have considered one-way random-effect models.

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$\alpha_i \sim NID(0,\sigma^2_\alpha)$; $\epsilon_{ij} \sim NID(0,\sigma^2)$; all α_i and ϵ_{ij} mutually independent. The likelihood function is

$$L(\sigma^2,\sigma^2_\alpha,\mu/y) \;\alpha\; |V|^{-\frac{1}{2}} Y'V^{-1}Y$$

where

$$Y = \begin{bmatrix} y_{11}-\mu \\ y_{12}-\mu \\ \cdots \\ y_{1n}-\mu \\ \vdots \\ y_{kn}-\mu \end{bmatrix} ; \quad \underset{(kn \times kn)}{|V|} = (A/0) \quad ; \quad \underset{(n \times n)}{A} = ((\sigma^2+\sigma^2_\alpha)/\sigma^2_\alpha)$$

and (a/b) denotes a square matrix whose all diagonal elements are a and all off-diagonal elements are b, the elements a and b may be square matrices of appropriate order.

Now

$$|A| = (\sigma^2+n\sigma^2_\alpha)(\sigma^2)^{n-1}$$

and, therefore

$$|V| = (\sigma^2+n\sigma^2_\alpha)^k(\sigma^2)^{k(n-1)}$$

and

$$V^{-1} = (B/0)$$

where

$$B = \left(\frac{\sigma^2+(n-1)\sigma^2_\alpha}{\sigma^2(\sigma^2+n\sigma^2_\alpha)} \Big/ -\frac{\sigma^2_\alpha}{\sigma^2(\sigma^2+n\sigma^2_\alpha)}\right) .$$

Thus

$$Y'V^{-1}Y = \frac{\sigma^2+(n-1)\sigma^2_\alpha}{\sigma^2(\sigma^2+n\sigma^2_\alpha)} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij}-\mu)^2$$

$$- \frac{\sigma_\alpha^2}{\sigma^2(\sigma^2+n\sigma_\alpha^2)} \sum_{i=1}^{k} \sum_{j\neq j'}^{n} (y_{ij}-\mu)(y_{ij'}-\mu)$$

$$= \frac{1}{\sigma^2} \sum_i \sum_j (y_{ij}-\mu)^2 - \frac{\sigma_\alpha^2}{\sigma^2(\sigma^2+n\sigma_\alpha^2)} \sum_{i=1}^{k} [\sum_{j=1}^{n} (y_{ij}-\mu)]^2 .$$

Writing

$$y_{ij} - \mu = (y_{ij}-\bar{y}_{i.}) + (\bar{y}_{i.}-\bar{y}_{..}) + (\bar{y}_{..}-\mu)$$

and

$$\sum_j (y_{ij}-\mu) = n(\bar{y}_{i.}-\bar{y}_{..}) + n(\bar{y}_{..}-\mu)$$

where

$$\bar{y}_{i.} = \sum_{j=1}^{n} y_{ij}/n \quad ; \quad \bar{y}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}/kn \quad .$$

We have

$$Y'V^{-1}Y = (\frac{1}{\sigma^2}) \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij}-\bar{y}_{i.}) + (\frac{1}{\sigma^2+n\sigma_\alpha^2})[n \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + nk(\bar{y}_{..} - \mu)^2] .$$

Denoting

$$S_1 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij}-\bar{y}_{i.})^2 \quad ; \quad S_2 = n \sum_{i=1}^{k} (\bar{y}_{i.}-\bar{y}_{..})^2$$

we have

$$L(\sigma^2,\sigma_\alpha^2,\mu/y) \alpha (\sigma^2)^{-\frac{k(n-1)}{2}} (\sigma^2+n\sigma_\alpha^2)^{-\frac{k}{2}} \times \mathrm{Exp}[-\tfrac{1}{2}\{\frac{S_1}{\sigma^2} + \frac{S_2+kn(\bar{y}_{..}-\mu)^2}{\sigma^2+n\sigma_\alpha^2}\}]$$

$$-\infty < \mu < \infty ; \ \sigma^2 \geq 0 \ ; \ \sigma_\alpha^2 \geq 0 \quad .$$

As we are interested in the component of variance σ^2 and σ_α^2, we eliminate μ by maximizing the likelihood function with respect to μ and obtain

$$\underset{\max/\mu}{L}(\sigma^2,\sigma_\alpha^2/y)\alpha(\sigma^2)^{-\frac{k(n-1)}{2}}(\sigma^2+n\sigma_\alpha^2)^{-\frac{k}{2}}\operatorname{Exp}[-\tfrac{1}{2}\{\frac{S_1}{\sigma^2}+\frac{S_2}{\sigma^2+n\sigma_\alpha^2}\}]$$

$$\sigma^2 \geq 0\ ;\ \sigma_\alpha^2 \geq 0\ .$$

The maximum likelihood estimates of σ^2 and σ_α^2 are

$$\hat{\sigma}^2 = \frac{S_1}{k(n-1)}$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{n}[\frac{S_2}{k} - \frac{S_1}{k(n-1)}] \quad \text{if} \quad S_1 < (n-1)S_2$$

and

$$\hat{\sigma}^2 = \frac{S_1 + S_2}{kn}$$

$$\hat{\sigma}_\alpha^2 = 0 \qquad\qquad \text{if} \quad S_1 > (n-1)\ S_2\ .$$

It may be mentioned that Thompson (32) has used the location invariant part of the sufficient statistics $(\bar{y}_{..},S_1,S_2)$ i.e. (S_1,S_2) to obtain the following likelihood function of σ^2 and σ_α^2,

$$L(\sigma^2,\sigma_\alpha^2/S_1,S_2)\alpha(\sigma^2)^{-\frac{k(n-1)}{2}}(\sigma^2+n\sigma_\alpha^2)^{-\frac{(k-1)}{2}}\operatorname{Exp}[-\tfrac{1}{2}\{\frac{S_1}{\sigma^2}+\frac{S_2}{\sigma^2+n\sigma_\alpha^2}\}]$$

$$\sigma^2 \geq 0\ ;\ \sigma_\alpha^2 \geq 0\ .$$

The maximum likelihood estimates of σ^2 and σ_α^2 derived from this likelihood function are what Thompson calls "restricted maximum likelihood estimates". They are given below.

$$\hat{\sigma}^2 = \frac{S_1}{k(n-1)}$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{n}[\frac{S_2}{k-1} - \frac{S_1}{k(n-1)}] \quad \text{if} \quad S_1 < \frac{k(n-1)}{(k-1)}\ S_2$$

and

$$\hat{\sigma}^2 = \frac{S_1 + S_2}{kn-1}$$

$$\hat{\sigma}^2_\alpha = 0 \qquad \text{if} \quad S_1 \geq \frac{k(n-1)}{(k-1)} S_2 \quad .$$

It may be noted that for large values of k, we have the same likelihood function and the same maximum likelihood estimates of σ^2 and σ^2_α.

The likelihood principle does not involve any strict accept — reject rules as in the Neyman-Pearson theory of hypothesis testing. The evidence supporting one hypothesised value of the parameter against the other is given by relative values of the likelihood function i.e. the likelihood ratio. This is a simple comparison and does not involve the concept of a particular alternative. We can, therefore, make a rule to compare two competing hypotheses in such a way that the likelihood ratio is a number between 0 and 1. The critical ratio is defined as a number λ $(0 \leq \lambda \leq 1)$ such that if the likelihood ratio is equal to or greater than λ, then the experimenter is not prepared to favour one hypothesis against the other or to change his opinion. For $\lambda = 1$, the likelihood principle states that both the hypotheses are equally plausible or in the language of the theory of support advocated by Hacking, the support given by the evidence to one hypothesis is the same as given to the other. The concept is similar to the significance level at which the experimenter is prepared to make a test of significance.

The difficulty with the likelihood principle is that it does not, in general, admit a probability statement. The very acceptance of the likelihood principle implies that the relative likelihoods will be the sole guide in the evaluation of two competing hypotheses. The likelihood axiom states that the likelihood function, determined by the observed sample, represents fully the evidence about parameter values. It is, therefore, necessary that a clear picture of the behaviour of the likelihood function is available to the experimenter. In one parameter case, it is simple. We can plot the likelihood of θ on a graph paper. The complexity of the problem increases with the number of parameters. In the case of two parameters, we have a surface of likelihoods to deal with. As it is the relative likelihood that matters in the present study, the best way of studying the behaviour of the likelihood function is to draw contours of equal likelihood expressed in terms of a common measure and the maximum likelihood value is a good measure to use. For this purpose, we have used the chosen sets of data in Appendix I. We have drawn contures of 50, 70, 90, 95 and 99 percent of the maximum likelihood for each of the fifteen sets of data in Appendix I. This provides us with a wide range of combinations of cumulative F-values and likelihoods relative to the maximum. Please see Figures 1 through 15.

The use of the likelihood function for inferential purposes depends on the type of inference. In the present study our aim is to have a better understanding of $(\sigma^2, \sigma_\alpha^2)$ from the given sample observations. The maximum likelihood values of σ^2 and σ_α^2 are best supported by the evidence. The interest does not end here. One may like to know how far a particular pair $(\sigma^2, \sigma_\alpha^2)$ is consistent with the data. The general procedure for an experimenter will, then, be to specify the critical ratio at which he is prepared to make comparisons. Suppose that his critical ratio is 0.50. Is this critical ratio poor, good or excellent? It is admitted that there is no answer for the present. The likelihood principle is not popular among statisticians and obviously the question of an agreement about the critical ratio to be used in a particular field or in general does not arise. However, after a decision (not in the sense of decision theory) has been made about the critical ratio to be used, the next step is to examine the system of contours of equal likelihoods for the critical ratio and above and make an inference. This will be illustrated with examples from the chosen sets of data. The systems of contours of equal likelihood are given in Figures 1 through 15.

Small cumulative F-value When the cumulative F-value is 0.25 or 0.50, the 'between' mean square is less than the 'error' mean square for all values of (k,n) considered. The maximum likelihood estimate of σ^2_α is zero. If we look at Figures (1, 2), (6, 7) and (11, 12) for (k,n) pairs (4, 5), (10, 5) and (20, 5) respectively, we will observe that we have a system of contours which cuts the σ^2-axis in each case and for all critical ratio from 0.50 and above. This shows that the hypothesis that $\sigma^2_\alpha = 0$ is well supported by the evidence. The sample size affects the peak of the contour along σ^2_α -axis. For example, with a cumulative F-value 0.25 and the critical ratio 0.50, the peak value of σ^2_α is 10.5 for $(k,n) = (4,5)$, 6.5 for $(k,n) = (10,5)$ and 4.60 for $(k,n) = (20,5)$. Similarly the distance between the two points at which the contour at a particular critical ratio cuts the σ^2-axis decreases with the sample size. For the cumulative F-value 0.25 and the critical ratio 0.50, the extreme values of σ^2 with $\sigma^2_\alpha = 0$ are (60,128), (73,117) and (80,112) for $(k,n) = (4,5)$, (10,5) and (20,5) respectively. It is interesting to note this is in agreement with sample theory results. For the same cumulative

F-values, the confidence interval or the confidence region is expected to decrease in size with an increase in the sample size. The evidence lends very good support to the hypothesis that σ_α^2 is nearly zero.

Large cumulative F-values   The cumulative F-values 0.95 and 0.99 imply that the analysis of variance test at the corresponding level of significance shows that σ_α^2 is significantly different from zero. The maximum likelihood estimate of σ_α^2 is a positive quantity in each case. This is reflected by the systems of the contours of equi-likelihood. A look at Figures (4, 5), (9, 10) and (14, 15) will reveal that the contours of critical ratio 0.50 and above are closed with the centre in the $(\sigma^2, \sigma_\alpha^2)$ quadrant. The hypothesis that σ_α^2 is different from zero is very well supported by the evidence. The remarks regarding the peak of a contour along σ_α^2-axis and the difference between the two extreme values of σ^2 for a given σ_α^2 also apply in this case, being the property of sample size. It may be noted that if an experimenter is prepared to lower his critical ratio of 0.50 then an hypothesis with $\sigma_\alpha^2 = 0$ can have the same support as another hypothesis with σ_α^2 different from zero as we observe that the lower part 50 percent of maximum likelihood contour is close to $\sigma^2\sigma$axis and with a decrease in the contour value it may cut the σ^2-axis. The evidence lends an excellent support to the hypothesis that σ_α^2 is different from zero.

Moderate cumulative F-value   For this purpose we have taken 0.75 as the cumulative F-value. The usual analysis of variance test with the frequently used levels of significance at 0.95 and 0.99, will show that σ_α^2 is not significantly different from zero, but the maximum likelihood estimate of σ_α^2 is a positive quantity. This is reflected by the system of contours in Figures 3, 8 and 11. It can be observed that at higher critical ratio, the contours are closed in σ^2 , σ^2_α quadrant. The value of the least critical ratio at which the contour is closed decreases with the sample size, as one would expect from the sample theory point of view. The evidence well supports the hypothesis that σ^2_α is small but different from zero.

One may be interested in the ratio σ^2_α/σ^2 . A study of likelihood contours is profitable for this information. We have selected one open and one closed system of contour for each sample size (k,n). The procedure is to draw tangent lines to the contour of the cirtical ratio which pass through the origin. For the sample size (k,n) = (4,5), we have selected the cumulative F-values 0.50 and 0.99. With the critical ratio as 0.50, the interval for σ^2_α/σ^2 , which we can call 50 percent critical interval, is (0.0 - 0.20) when cumulative F-value is 0.50 and (0.02 - 0.20) when the cumulative F-value is 0.99. The concept of the critical interval is similar to confidence interval with the difference that no probability statement can be attached to the critical interval due to the very nature of the likelihood inference. For the sample size (k,n) = (10,5) the selected values of cumulative F are 0.75 and 0.95 and the corresponding 50 percent critical intervals are (0.0 - 0.25) and (0.02 - 0.52) respectively. For the moderately large sample size (k,n) = (20,5), the selected cumulative F- values are 0.25 and 0.99. The corresponding 50 percent critical intervals are (0 - 0.05) and (0.075 - 0.413).

An objection that can be raised against this type of inference is that at a given critical ratio, widely apart values of $(\sigma^2, \sigma^2_\alpha)$ have the same level of support. In fact all pairs on a contour of a given likelihood have the same level of support. The relative likelihood of two pairs on a given contour is 1 and according to the likelihood inference, both the pairs are equally plausible. There is no basis to discriminate one from the other. But this is not peculiar to the likelihood inference. The approximate confidence intervals discussed in Chapter II may be wide. The approximate confidence region for two parameters θ_1 and θ_2 proposed by Bartlett (3,4) is a contour in θ_1 , θ_2 plane. As we shall see in Chapter V, the contours of equal goodness of fit have the same defect, if it is a defect at all. This may be called the curse of dimensionality. With one restriction, the three-dimensional space $(\theta_1, \theta_2, f(\theta_1, \theta_2))$ is reduced to the two dimension space (θ_1, θ_2) and the proposed criterion, providing the restriction, gives a contour in (θ_1, θ_2) plane.

B. Likelihood and Posterior Distribution

The likelihood function plays an important role in the Bayesian analysis. The posterior distribution is proportional to the product of the likelihood and the prior used. When we change the prior distribution, the likelihood function remains the same and the posterior distribution changes. The prior may change from person to person and for the same person from time to time. This has led some Bayesian statisticians to think that it is the likelihood function that matters and the job of a statistician is to report the likelihood function to the experimenter. He may use his own prior and get his posterior distribution to draw inference about the parameter. It is, therefore, important to know how a prior enters the likelihood function to give a posterior distribution.

We shall consider the non-informative priors (i) $\frac{d\sigma^2}{\sigma^2} \cdot \frac{d\tau^2}{\tau^2}$, (ii) $\frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau}$ with $\tau^2 = \sigma^2 + \sigma_\alpha^2$ as used by us and $\tau^2 = \sigma^2 + n\sigma_\alpha^2$ as used by others for each prior. For the purpose of illustration, we use the sets of data in Appendix A. We have considered two sample sizes viz $(k,n)=(4,5)$ and $(k,n) = (20.5)$, representing small and moderately large sample sizes respectively. For each selected sample size, we have taken the sets of data with cumulative F-value 0.50 and 0.95. This will give us one open and one closed system of contours. The contours of the likelihood function and the posterior distribution at 50 and 70 percent of their respective maximum with two contour of the prior for some chosen values of the prior have been drawn for each case; and are given in Figures 16 through 31.

(i) Prior $\alpha[\sigma^2(\sigma^2+\sigma_\alpha^2)]^{-1}$. This is the prior used by us in the present study. The prior is not completely absorbed in the likelihood function. The prior changes the exponent of σ^2 from $-[\frac{k(n-1)}{2}]$ to $-[\frac{k(n-1)+2}{2}]$. The term $(\sigma^2+\sigma_\alpha^2)^{-1}$ in the prior is not absorbed into likelihood, but the prior has an effect on the σ^2- component of the mode of the posterior distribution, which is less than the corresponding component in the mode of the likelihood function. A similar effect is observed in σ_α^2 - component of the mode of the posterior distribution, except when S_1 and S_2 are such that the σ_α^2 - component of the mode is taken as zero for both the likelihood function and the posterior distribution. A study of the Figures 16, 20, 24, and 28 will show that the contours of the likelihood function and the posterior distribution at the same level of their respective maximum likelihood have more or less the same shape but the size, as measured by the area covered, is small for the contours of the posterior distribution. A shift towards the origin along both the axes is observed indicating the effect of σ^2 and $(\sigma^2+\sigma_\alpha^2)$ in the prior. The points $(\sigma^2,\sigma_\alpha^2)$ are not so widely apart on a given contour of the posterior as they were on the corresponding contour of the likelihood. According to the likelihood inference, an hypothesized pair $(\sigma^2,\sigma_\alpha^2)$ which was acceptable at a given critical ratio may not now be acceptable at the same critical ratio. The effect of prior on the likelihood is more marked when the sample size is small. Similar is the case with 50 and 70 percent (of the maximum likelihood) contours. The curve $[\sigma^2(\sigma^2+\sigma_\alpha^2)]^{-1}$ = constant is a branch of the hyperbola $x(x+y)$ = constant. This is illustrated in Figures 16, 20, 24, and 28.

(ii) Prior $\alpha[\sigma^2(\sigma^2+n\sigma_\alpha^2)]^{-1}$. As observed earlier, the advantage of working with this prior is that the prior is completely absorbed in the likelihood function. We can consider the posterior distribution as the likelihood function of $(\sigma^2,\sigma_\alpha^2)$ with the same S_1 and S_2 values, but with degrees of freedom for 'between' mean square and error mean square increased by 2 in each case. The σ^2 and σ_α^2 components of the mode of the posterior distribution are less than the corresponding components of the mode of the likelihood. A study of Figures 17, 21, 25 and 29 will show that the remarks on the size of contours, the likelihood inference etc., given at (i) above, are equally applicable in this case, also. Due to the factor n in the prior under discussion, the peak of a contour along σ_α^2 - axis is shifted more toward the origin than the peak of the contour of the same value with prior at (i) above. The curve $[\sigma^2(\sigma^2+n\sigma_\alpha^2)]^{-1}$ = constant is a branch of the hyperbola $x(x+ny)$ = constant. The hyperbola with this prior has less slope than the corresponding hyperbola with the prior at (i) above.

(iii) Prior $\alpha(\sigma^2+\sigma_\alpha^2)^{-1}$. We have also used this prior in the present study. This is not absorbed in the likelihood function. A point $(\sigma^2,\sigma_\alpha^2)$ receives the weight in proposition to the inverse of sum of its coordinates in $(\sigma^2,\sigma_\alpha^2)$ plane. The effect of this prior on the likelihood function is the same in nature as the effect of the prior $\alpha[\sigma^2(\sigma^2+\sigma_\alpha^2)]^{-1}$ but it is reduced in degree, perhaps due to the absence of the factor σ^2. A comparative study of Figures 20 and 22 is rewarding in this connection. The rest of the remarks regarding size of contours, the likelihood inference based on a given critical ratio etc., are the same in nature and different in degree and therefore, need not be repeated. The curve $(\sigma^2+\sigma_\alpha^2)^{-1}$ = constant is a straight line with a slope of 135^0. The results are illustrated in Figures 18, 22, 26 and 30.

(iv) Prior $\alpha(\sigma^2+n\sigma_\alpha^2)^{-1}$. This is the prior suggested by Stone and Springer (31). It is evident that the prior can be easily absorbed in the likelihood function. The posterior distribution can be considered as the likelihood function of σ^2 and σ_α^2 arising for a set of data with the same S_1, S_2 and degrees of freedom for the error mean square but with the degrees of freedom for 'between' mean square increased by 2. The prior affects σ_α^2 - component of the mode with the result that the center of contours of equi-likelihoods is shifted toward the origin along σ_α^2 - axis. The contour of the posterior for a given α percent of its own maximum likelihood is more flat than the corresponding contour of the likelihood function and the size is reduced. The effect of the prior on the likelihood inference based on a given critical ratio is the same in nature as discussed above. The curve $(\sigma^2+n\sigma_\alpha^2)^{-1}$ = constant is a straight line with a slope of $168^0 41'$ degrees. This case is illustrated in Figures 19, 23, 27, and 31.

Figure 1. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 4$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	3	122	40.7	0.41
Within	16	1600	100.0	

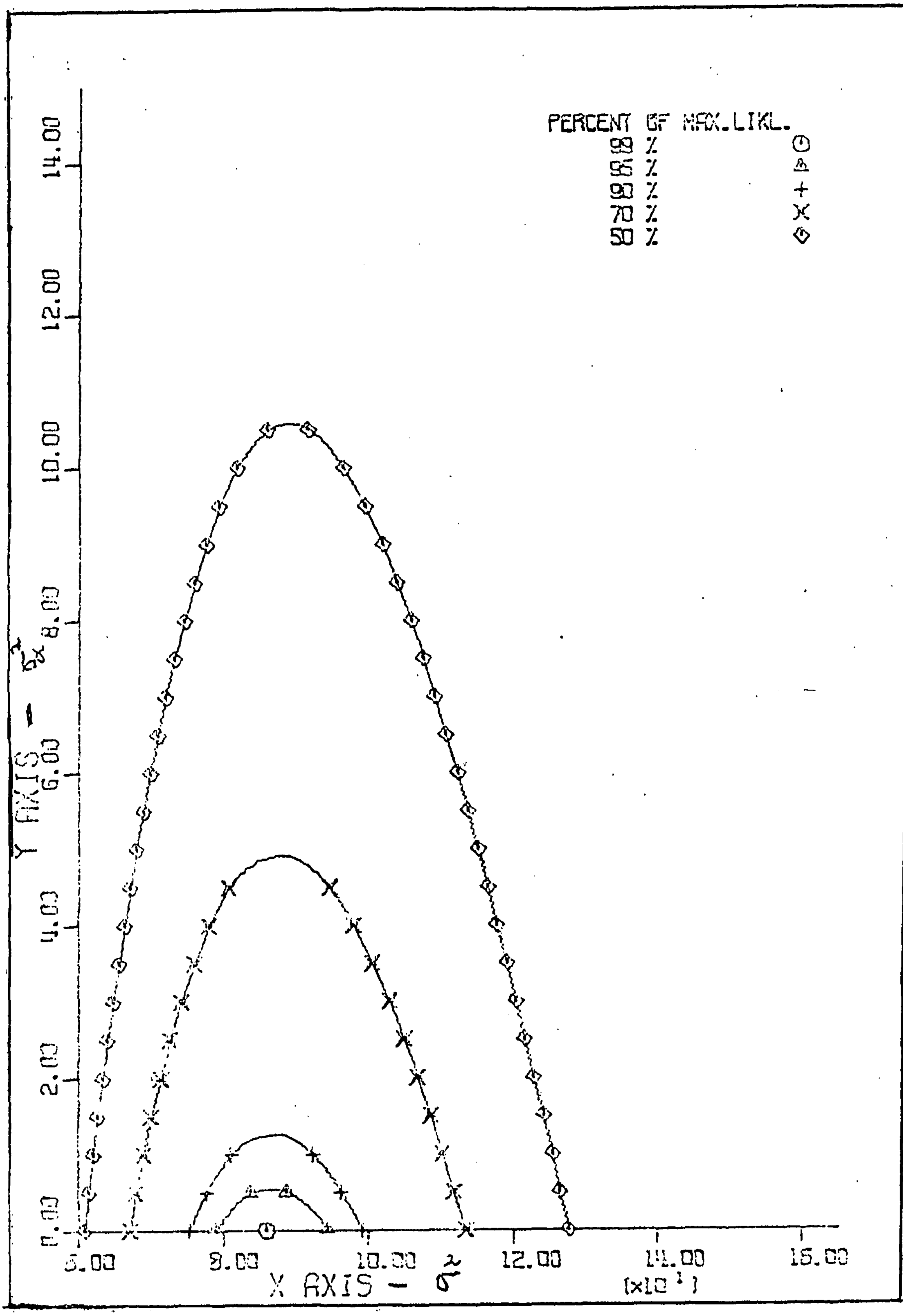

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS
X AXIS
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
6.00
8.00
10.00
12.00
14.00
16.00
(x10 1)

Figure 2. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 4$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	3	248	82.7	0.83
Within	16	1600	100.0	

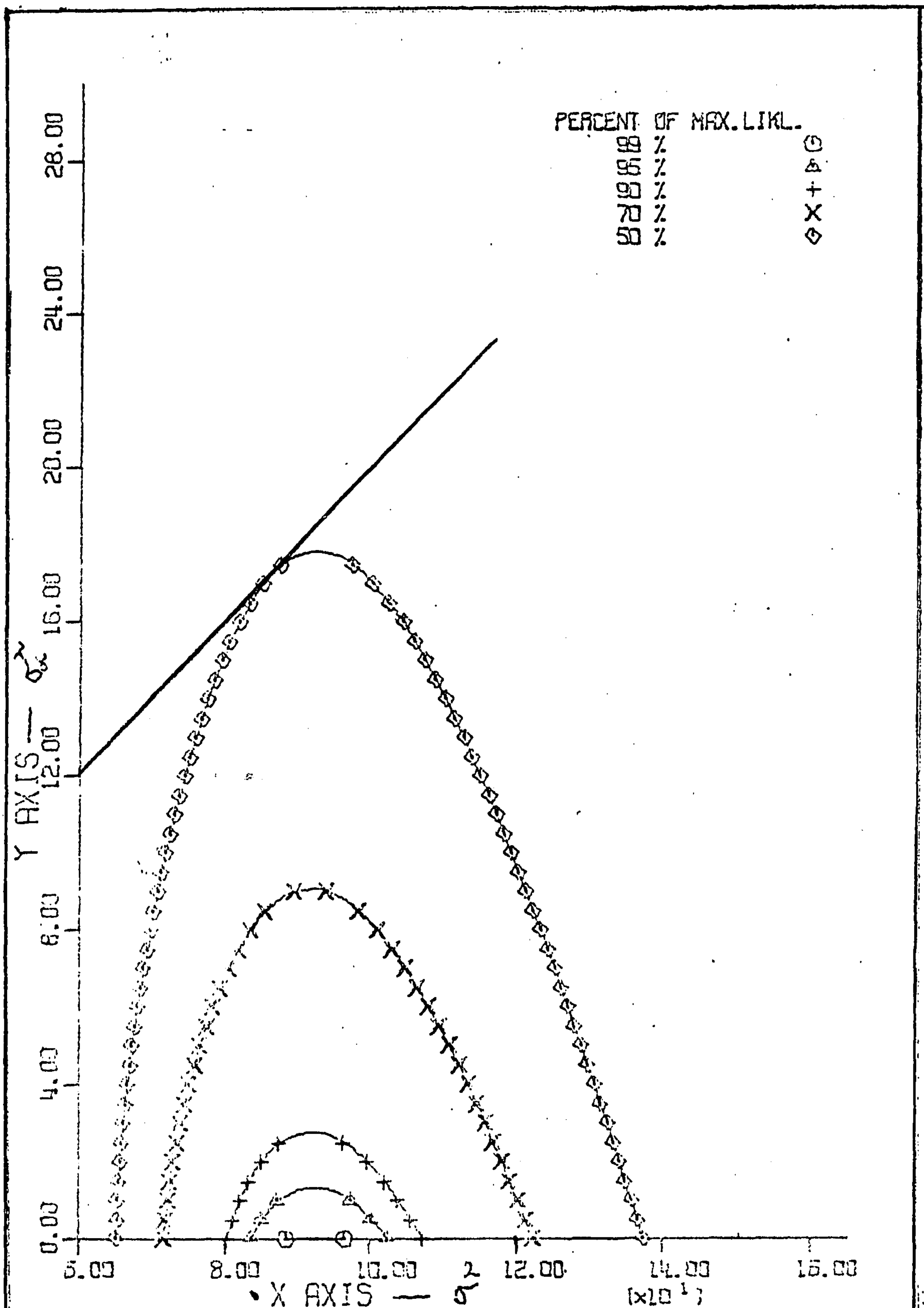
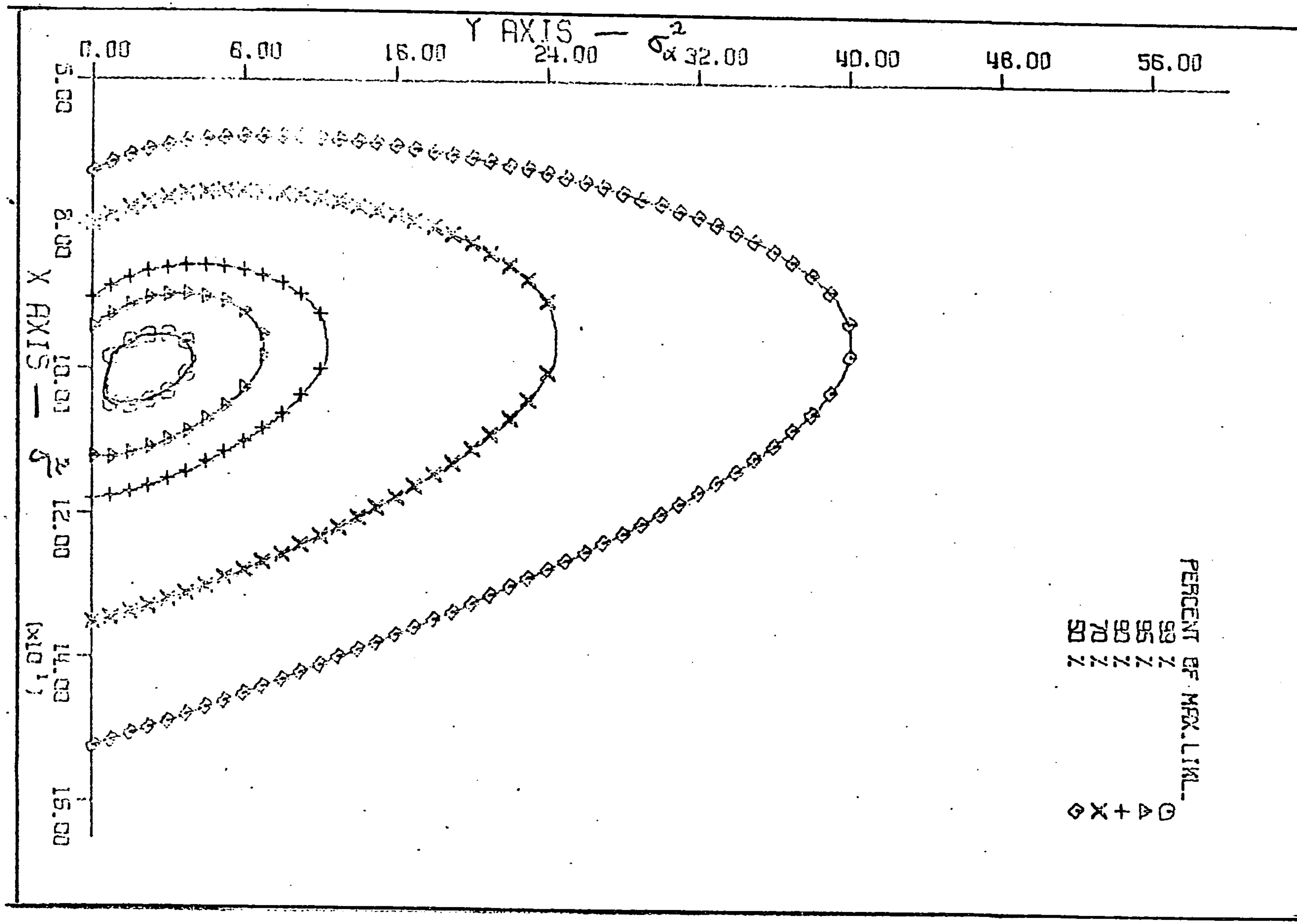

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — $\hat{\sigma}_{\alpha}^{2}$
0.00
4.00
8.00
12.00
16.00
20.00
24.00
28.00
6.00
8.00
10.00
12.00
14.00
16.00
[x10 1]
X AXIS — σ^{2}

Figure 3. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 4$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	3	454	151.3	1.51
Within	16	1600	100.0	



Y AXIS — σ_α^2
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
X AXIS — σ^2
6.00
8.00
10.00
12.00
14.00
16.00
(x10 1)
PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %

Figure 4. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 4$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	3	972	324.0	3.24
Within	16	1600	100.0	

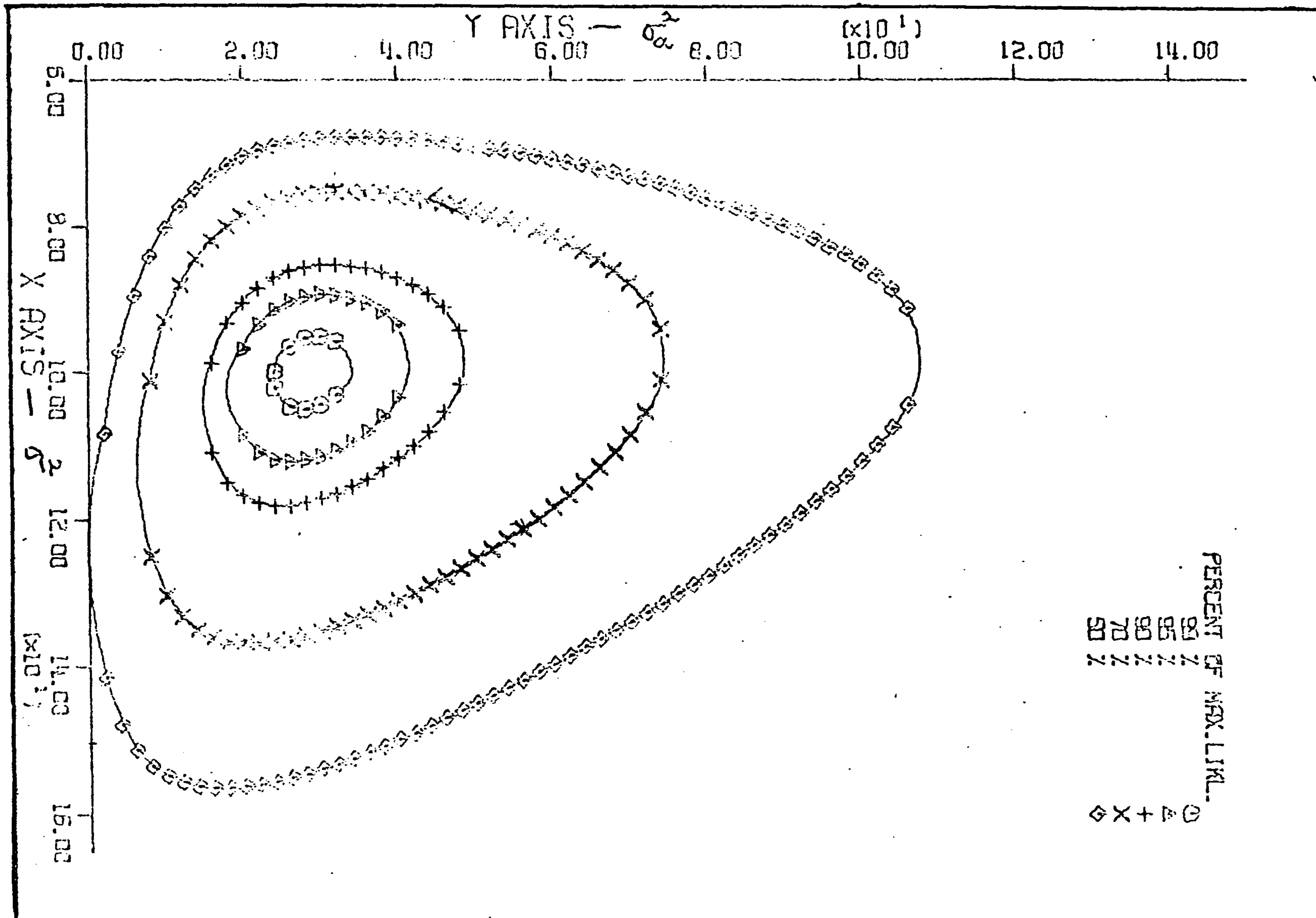

Y AXIS — σ_α^2
(x10 1)
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
X AXIS — σ^2
6.00
8.00
10.00
12.00
14.00
16.00
(x10 1)
PERCENT OF MAX.LIKL.

Figure 5. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 4$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	3	1588	529.3	5.29
Within	16	1600	100.0	

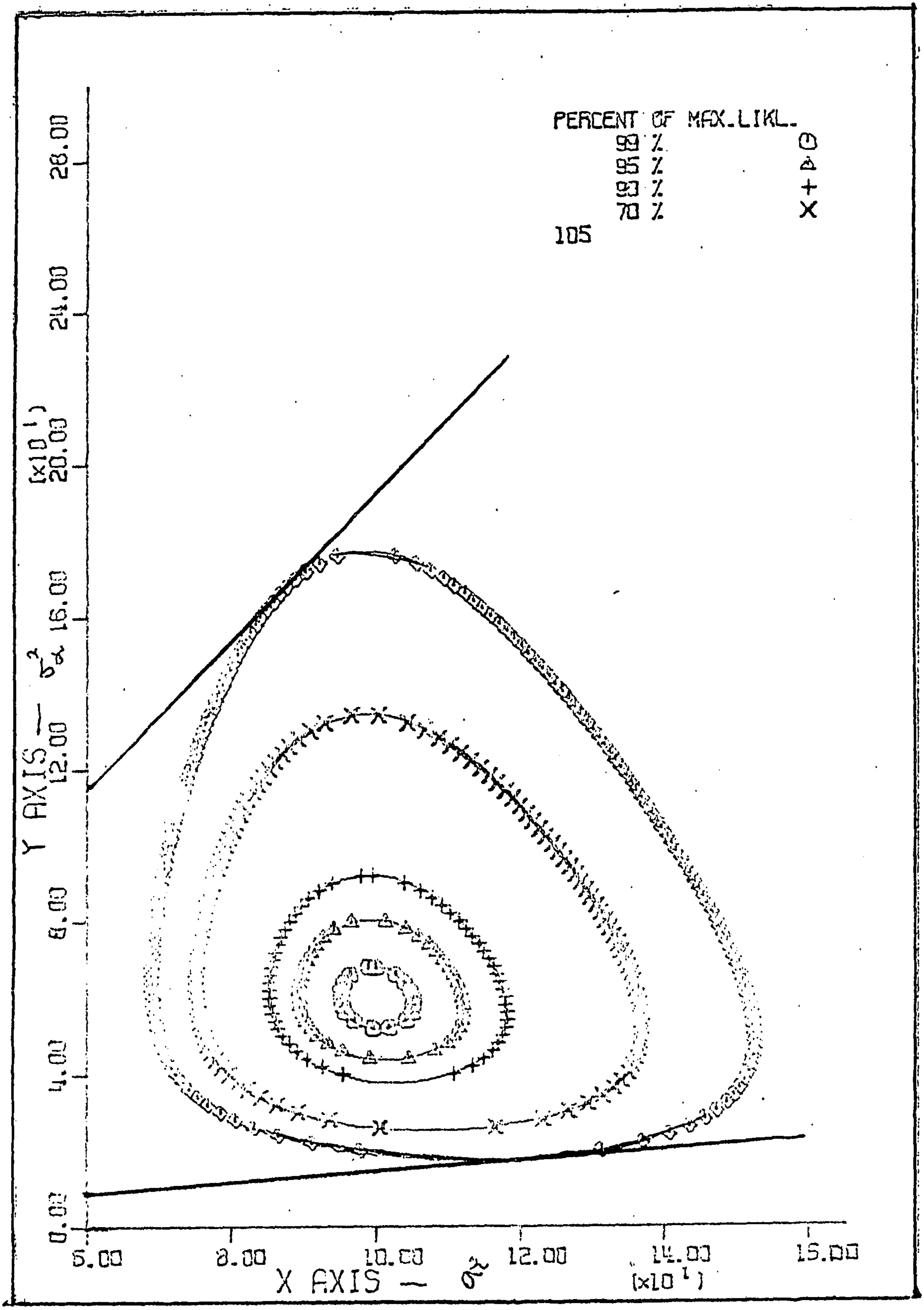

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
105
Y AXIS — σ_α^2
[x10 1]
0.00
4.00
8.00
12.00
16.00
20.00
24.00
28.00
X AXIS — σ_e^2
[x10 1]
6.00
8.00
10.00
12.00
14.00
15.00

Figure 6. Contours of equal likelihood as percent of the maximum likelihood for the data (k = 10, n = 5)

Source	d.f.	Sum of squares	M.S.S.	F
Between	9	582	64.7	0.65
Within	40	4000	100.0	

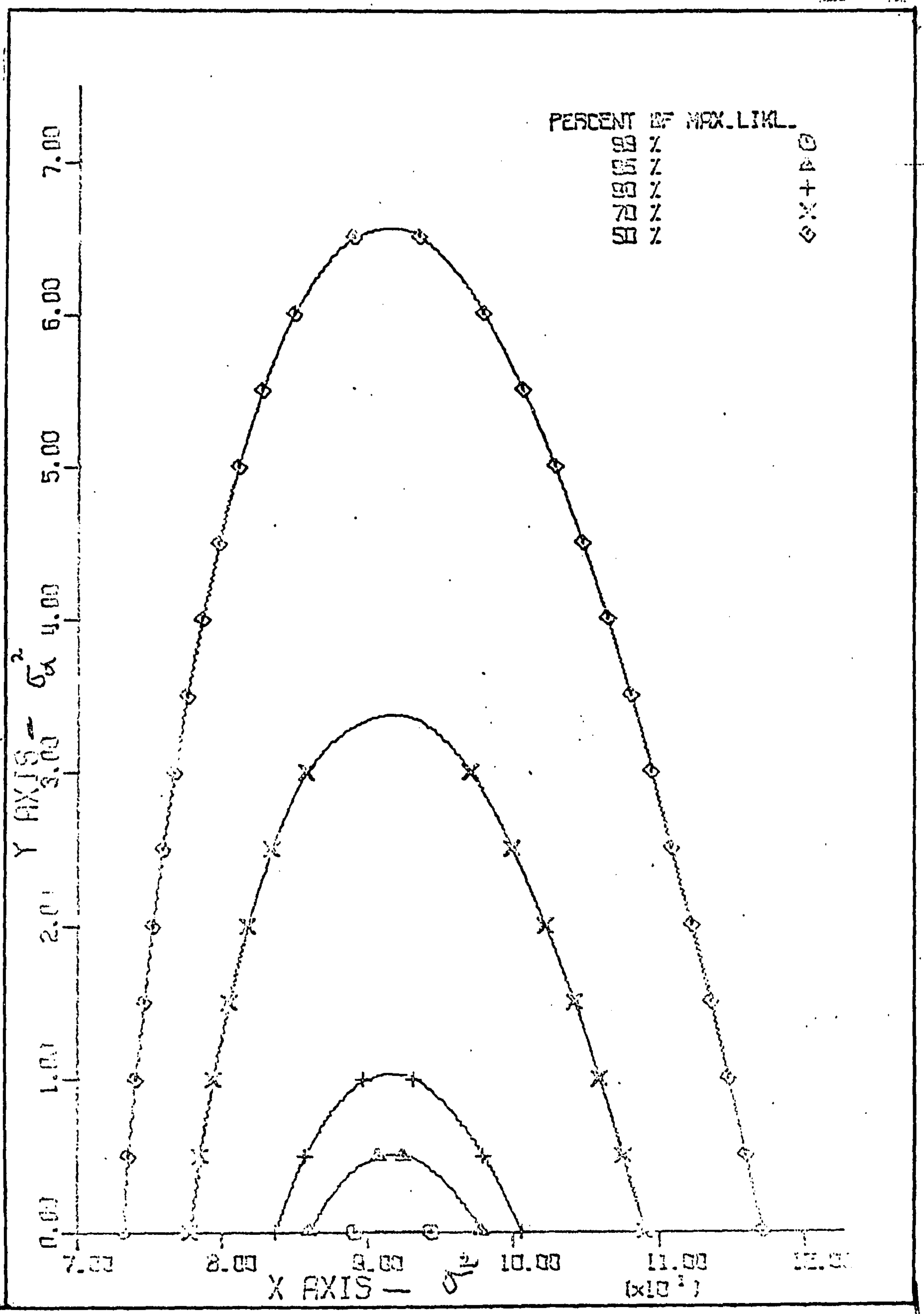

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — σ_α^2
X AXIS — σ^2
(x10^1)
7.00
8.00
9.00
10.00
11.00
12.00
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00

Figure 7. Contours of equal likelihood as percent of the maximum likelihood for the data (k = 10, n = 5)

Source	d.f.	Sum of squares	M.S.S.	F
Between	9	848	94.2	0.94
Within	40	41000	100.0	

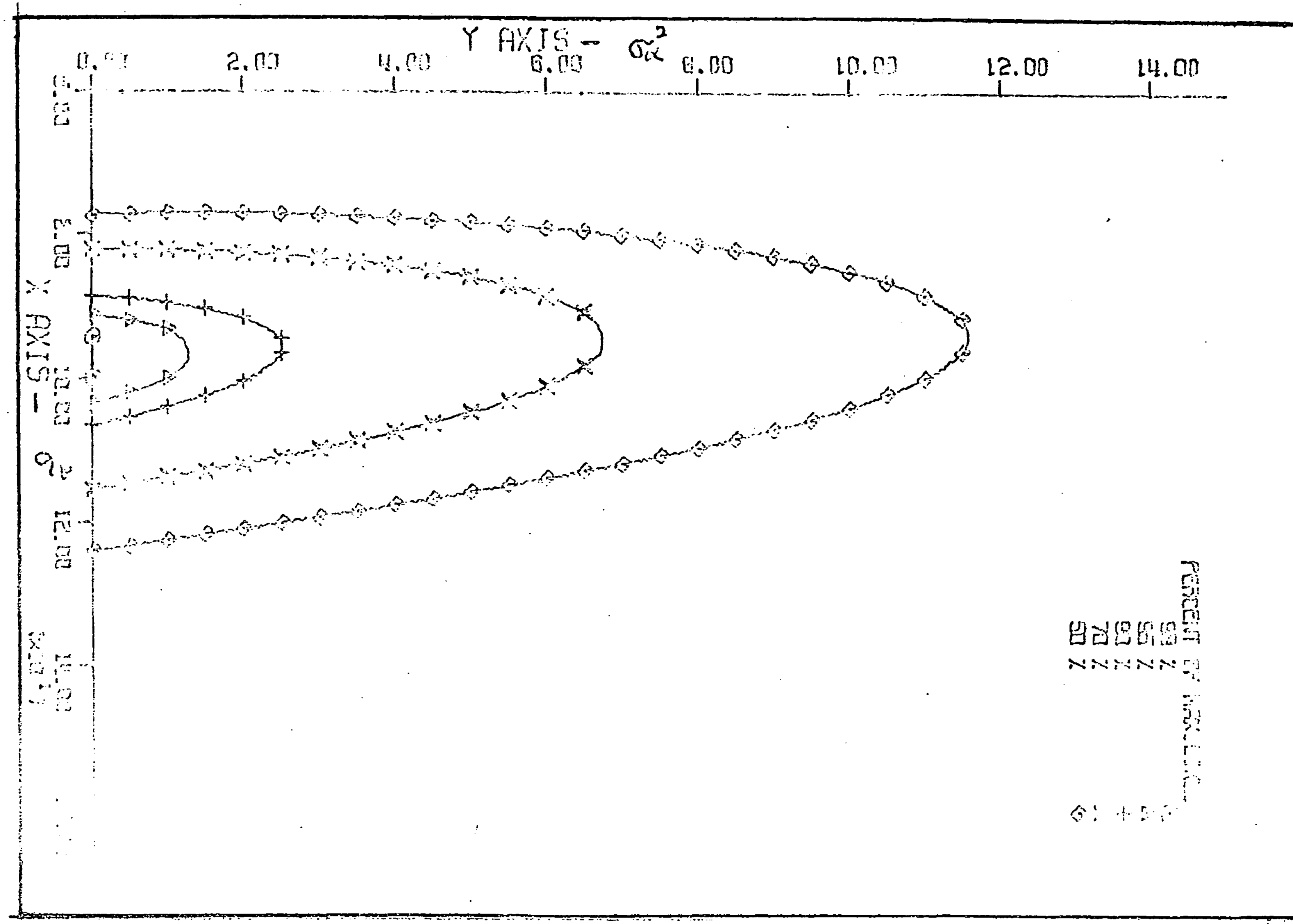

Y AXIS - σ_α^2
2.00
4.00
6.00
8.00
10.00
12.00
14.00
X AXIS -

Figure 8. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 10$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	9	1206	134.0	1.34
Within	40	4000	100.0	

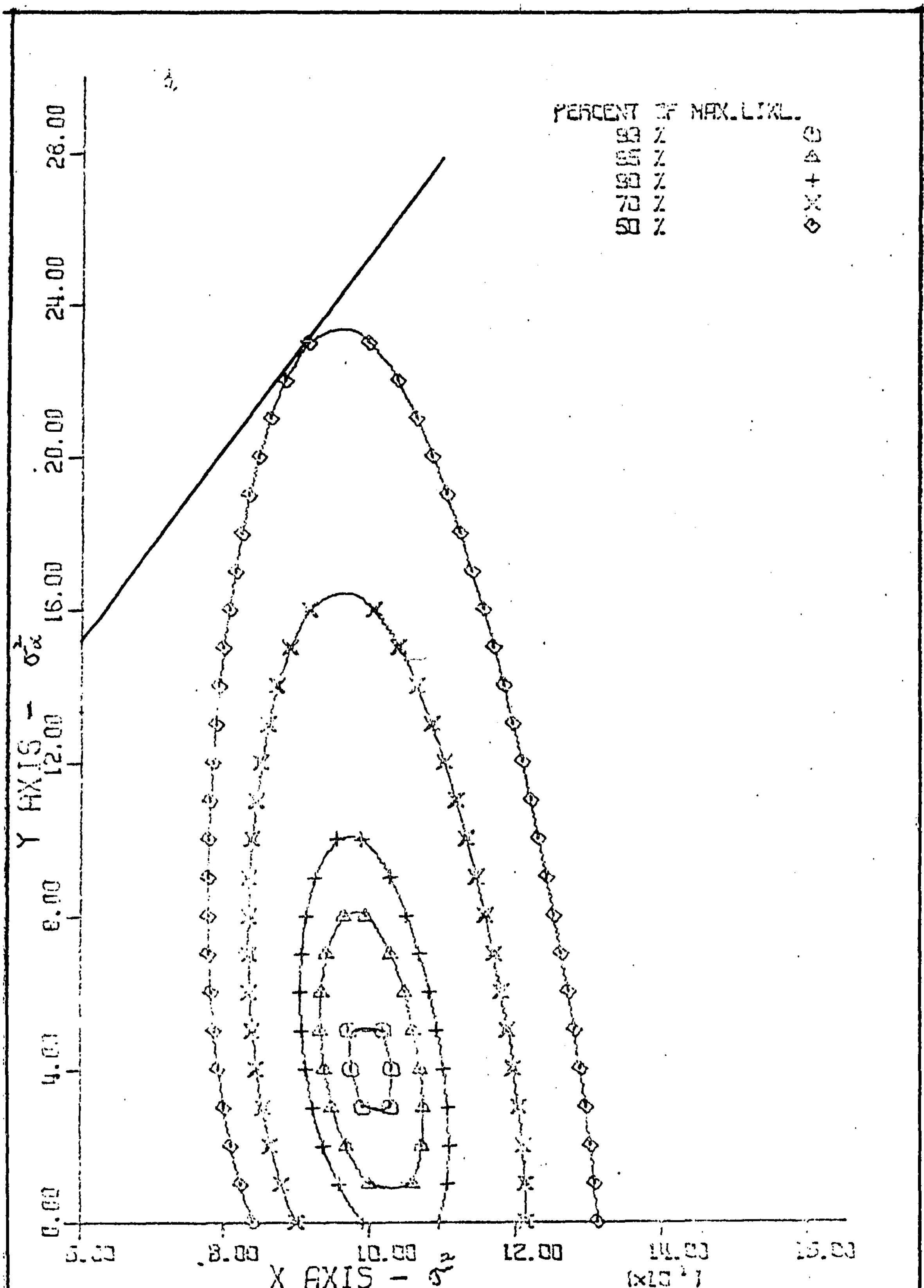

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS - σ_α^2
X AXIS - σ^2
0.00
4.00
8.00
12.00
16.00
20.00
24.00
28.00
8.00
10.00
12.00

Figure 9. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 10$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	9	1908	212.0	2.12
Within	40	4000	100.0	

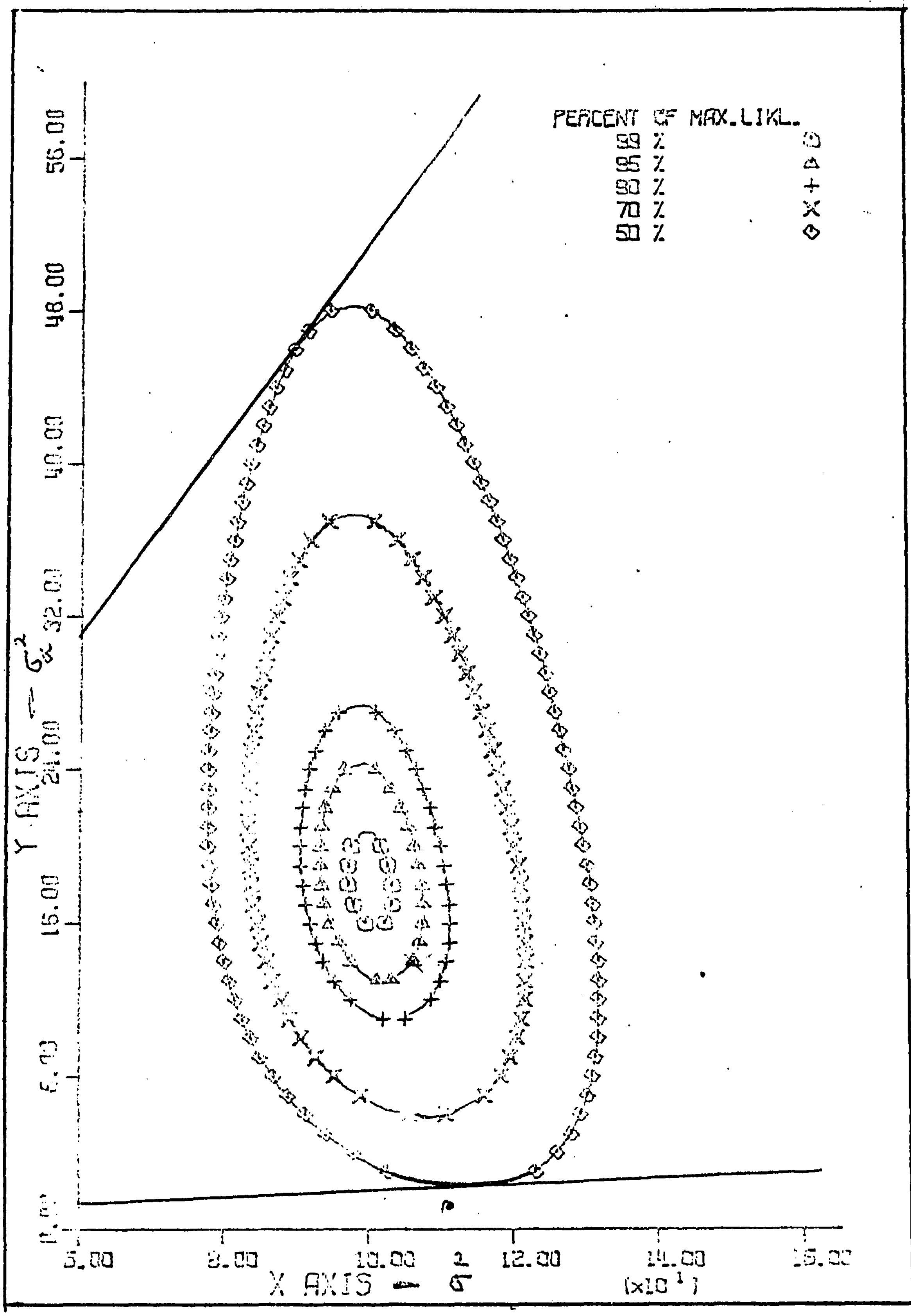

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — σ_x^2
X AXIS — $\frac{1}{\sigma}$
56.00
48.00
40.00
32.00
24.00
16.00
6.00
5.00
8.00
10.00
12.00
14.00
16.00
(x10 1)

Figure 10. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 10$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	9	2592	288.0	2.88
Within	40	4000	100.0	

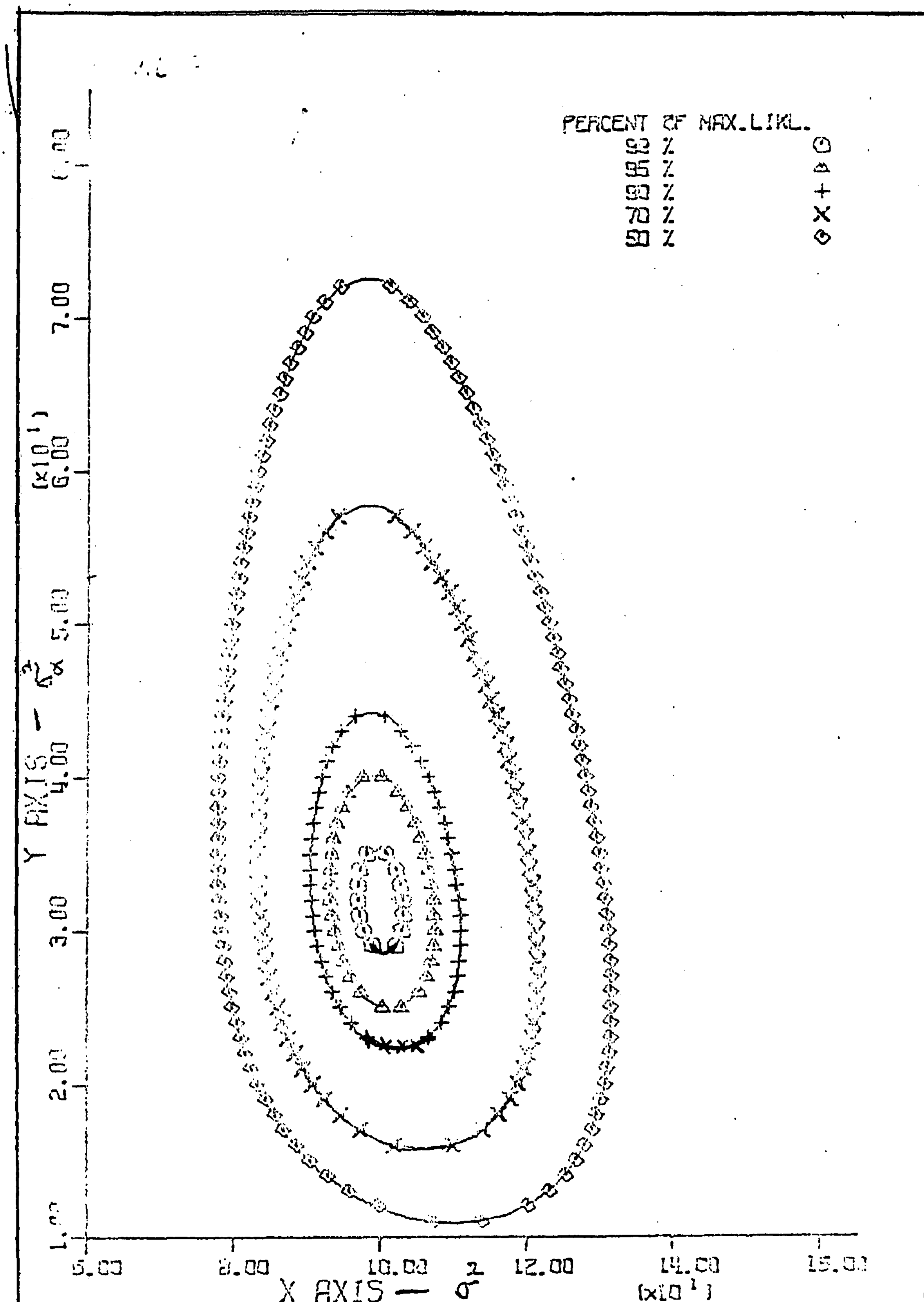

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — $\hat{\sigma}^2_\alpha$
(x10[1])
X AXIS — σ^2
(x10[1])
1.00
2.00
3.00
4.00
5.00
6.00
7.00
6.00
8.00
10.00
12.00
14.00
15.00

Figure 11. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 20$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	19	1434	75.5	0.76
Within	80	8000	100.0	

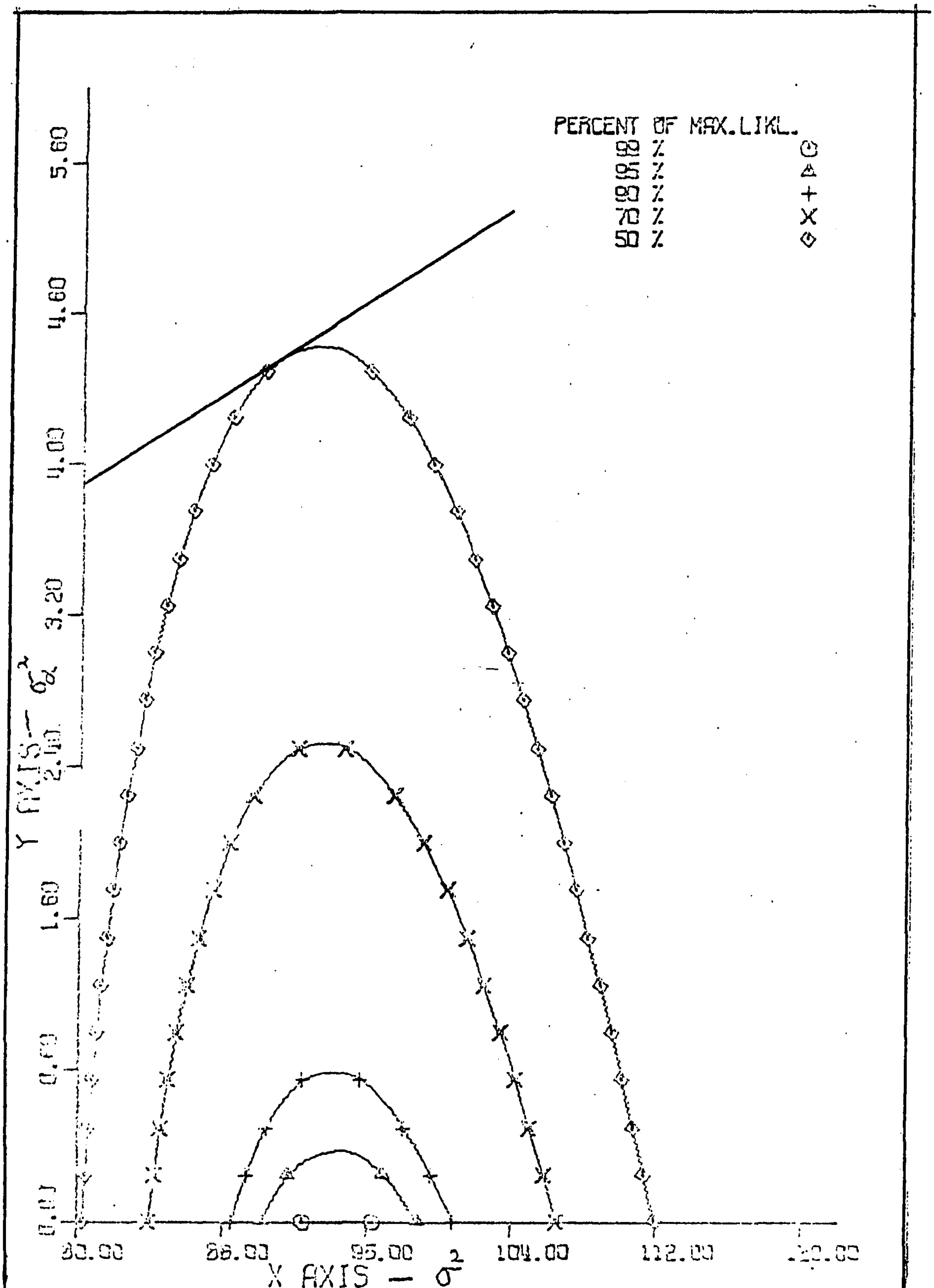

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — σ_α^2
0.00
0.80
1.60
2.40
3.20
4.00
4.80
5.60
X AXIS — σ^2
80.00
88.00
96.00
104.00
112.00

Figure 12. Contours of equal likelihood as percent of the maximum likelihood for the data (~~k~~ = 20, n = 5)

Source	d.f.	Sum of squares	M.S.S.	F
Between	19	1850	97.4	0.97
Within	80	8000	100.0	

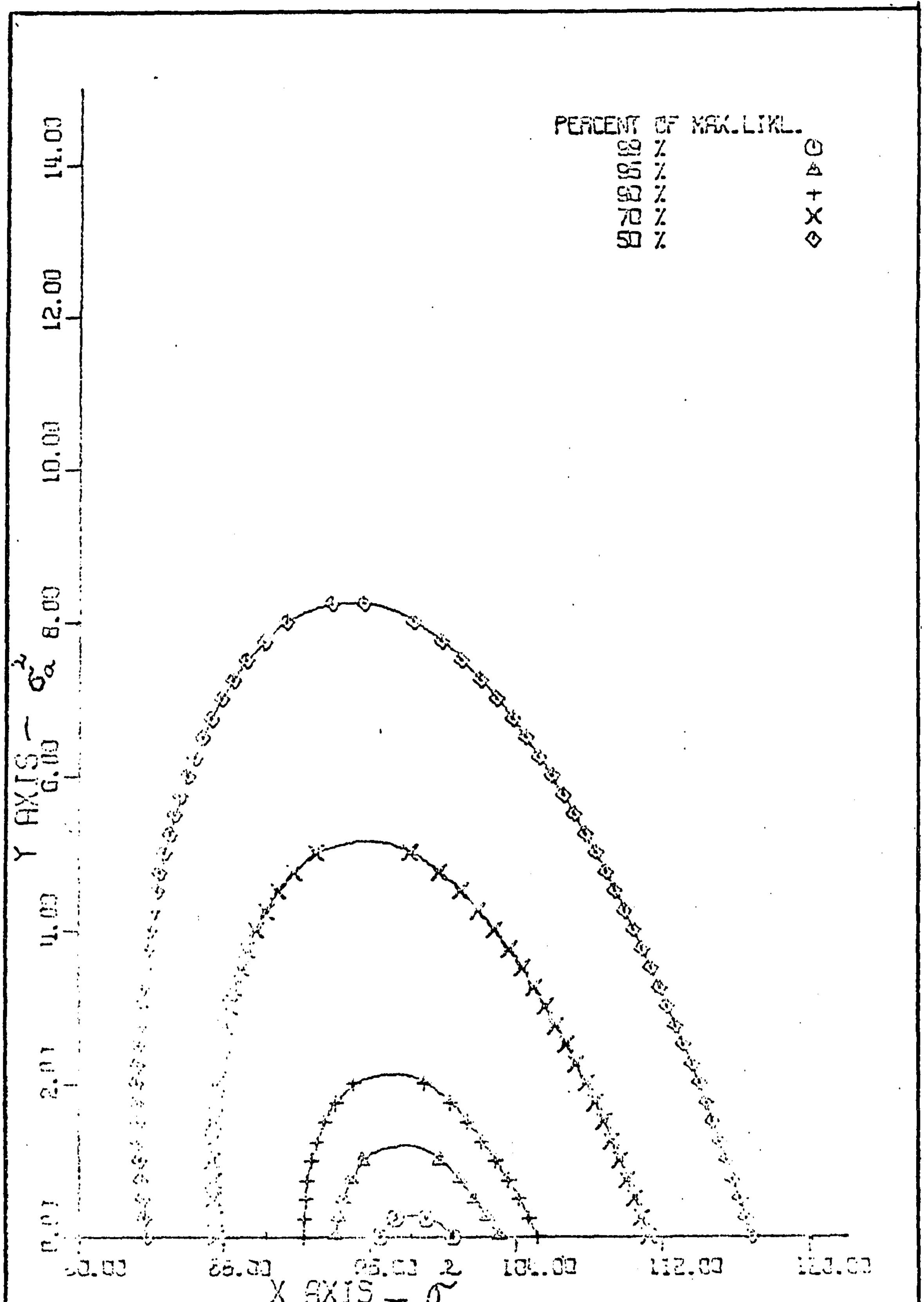

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — σ_α^2
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
80.00
88.00
96.00
104.00
112.00
120.00
X AXIS — σ^2

Figure 13. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 20$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	19	2356	124.0	1.24
Within	80	8000	100.0	

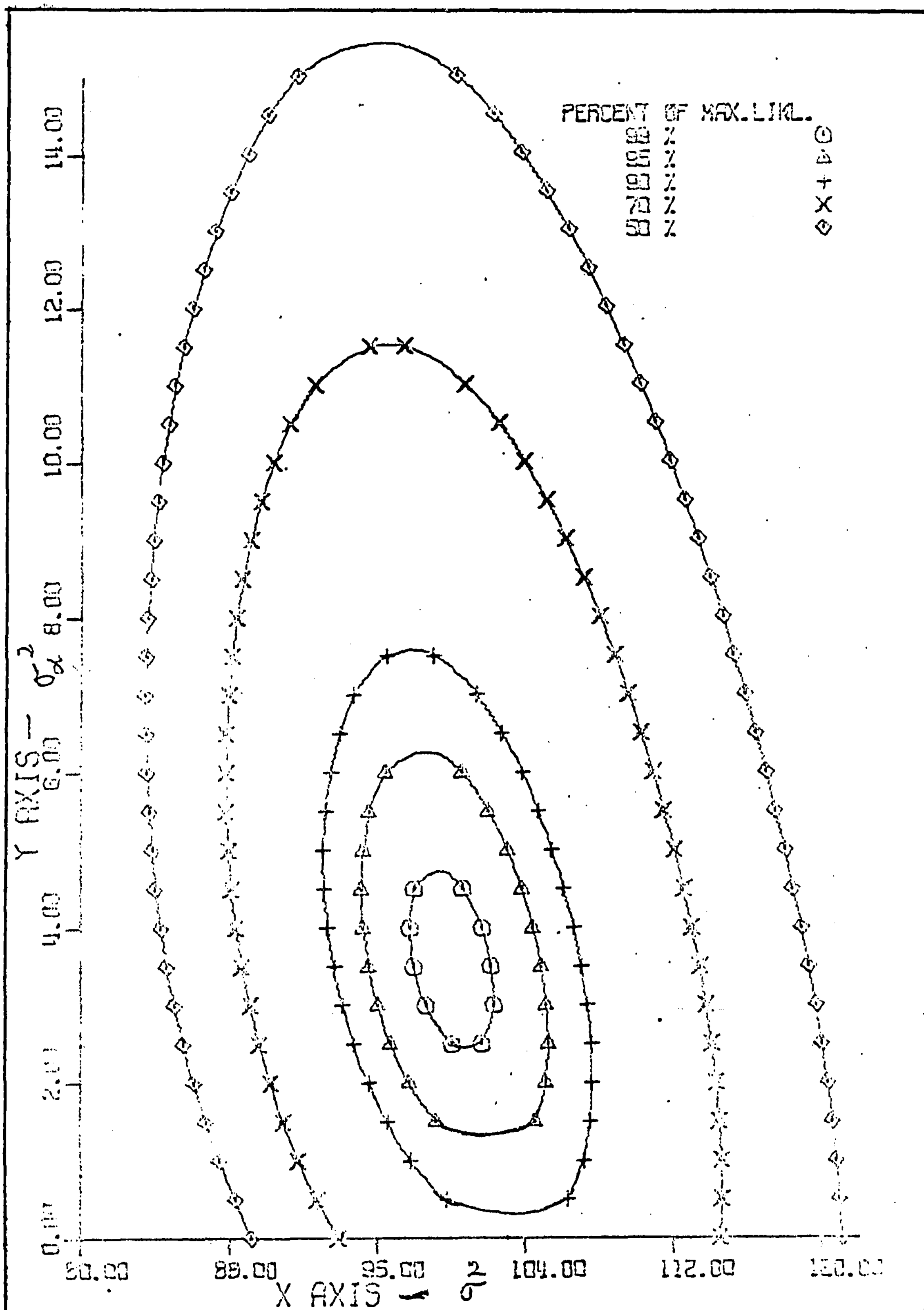

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — σ_α^2
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
X AXIS — σ^2
80.00
88.00
96.00
104.00
112.00
120.00

Figure 14. Contours of equal likelihood as percent of the maximum likelihood for the data (k = 20, n = 5)

Source	d.f.	Sum of squares	M.S.S.	F
Between	19	3268	172.0	1.72
Within	80	8000	100.0	

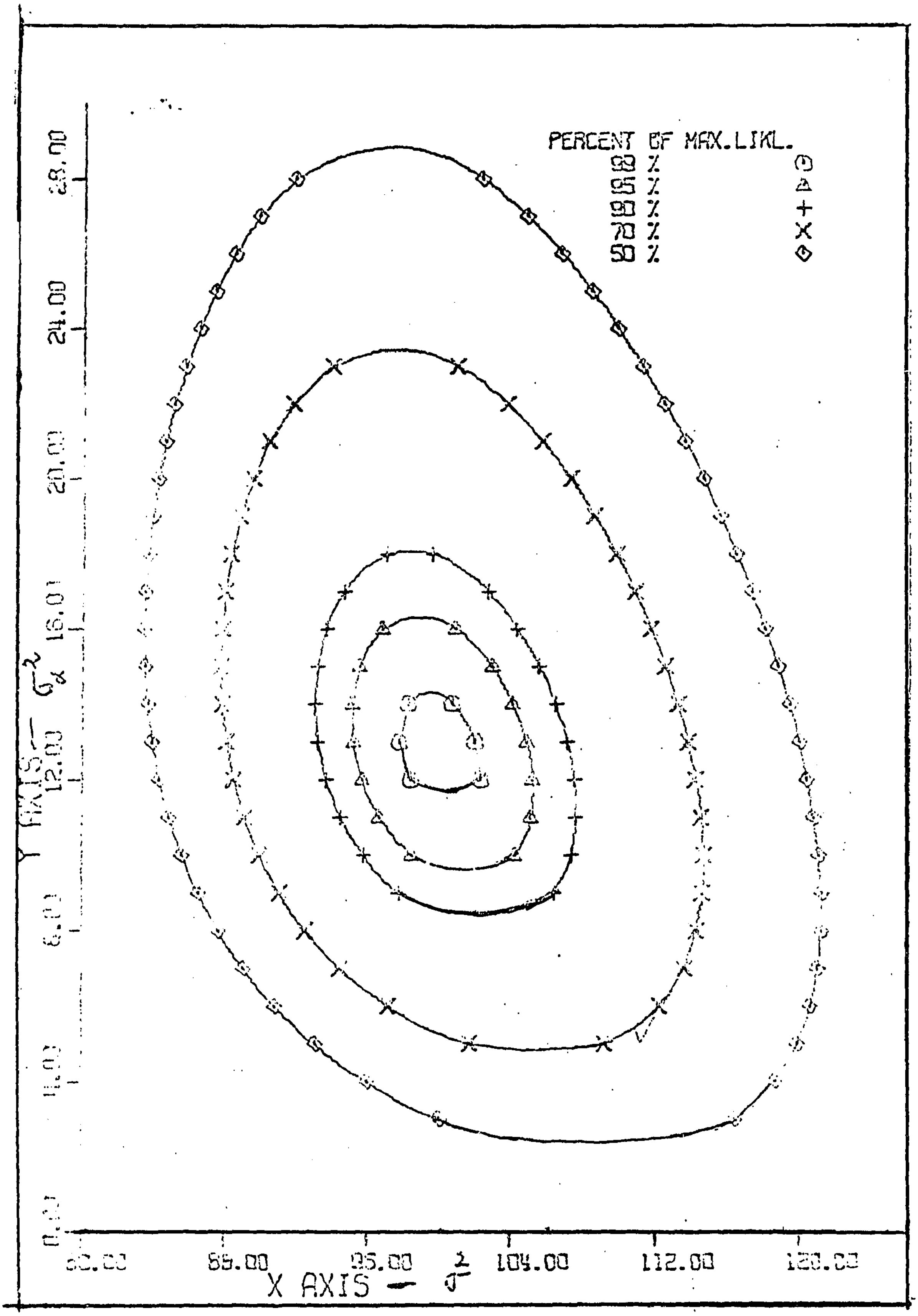

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
Y AXIS — σ_α^2
X AXIS — σ^2
104.00
112.00
24.00
28.00
20.00
16.00
12.00

Figure 15. Contours of equal likelihood as percent of the maximum likelihood for the data ($k = 20$, $n = 5$)

Source	d.f.	Sum of squares	M.S.S.	F
Between	19	4086	215.1	2.15
Within	80	8000	100.0	

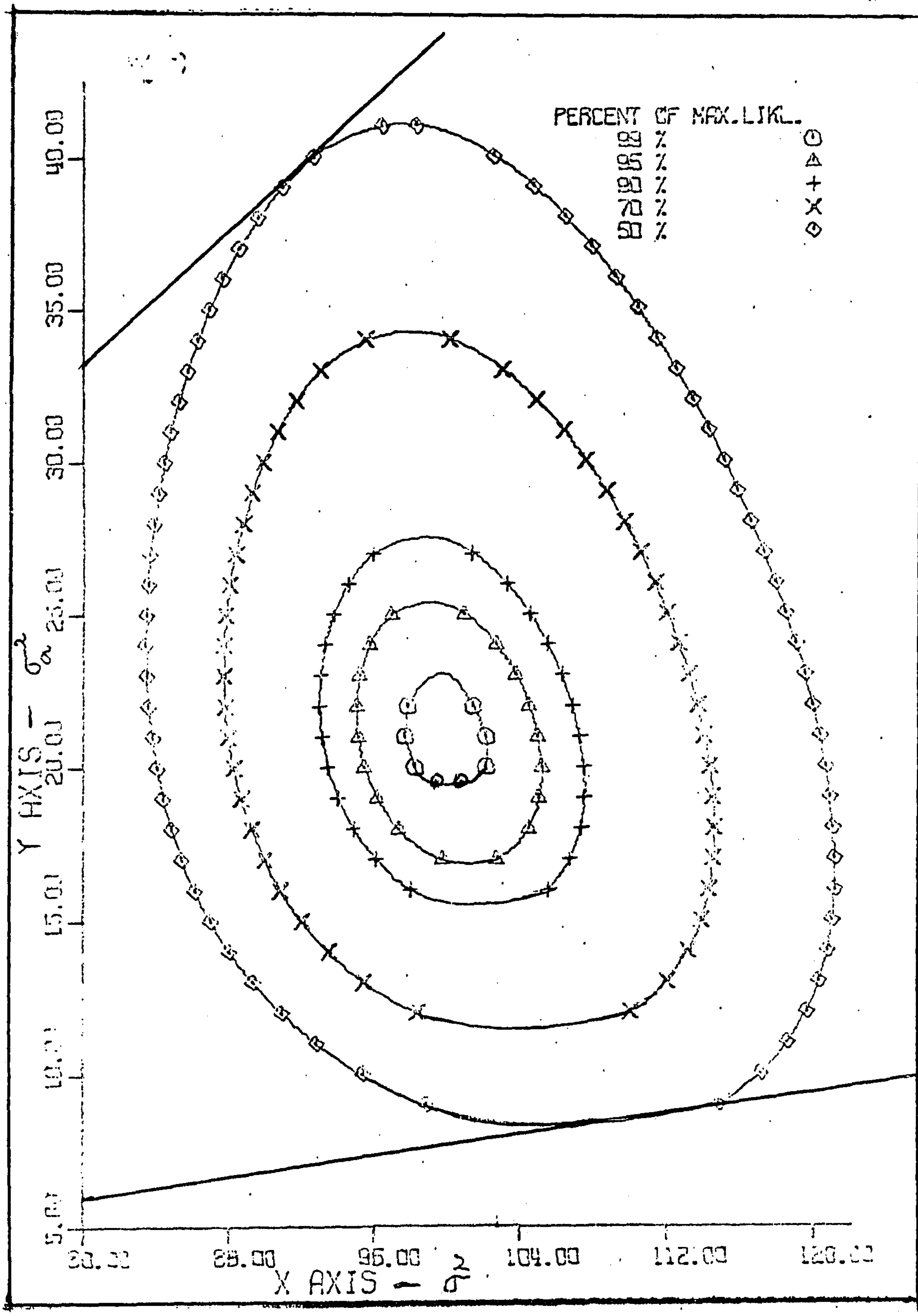

PERCENT OF MAX.LIKL.
99 %
95 %
90 %
70 %
50 %
40.00
35.00
30.00
25.00
20.00
15.00
Y AXIS — σ_α^2
88.00
96.00
104.00
112.00
120.00
X AXIS — σ^2

Figure 16. Contours of 50 percent of maximum for likelihood and posterior with prior $\alpha[\sigma^2(\sigma^2+\sigma_\alpha^2)]^{-1}$; $k = 4$, $n=5$; $S_1 = 1600$, $S_2 = 248$

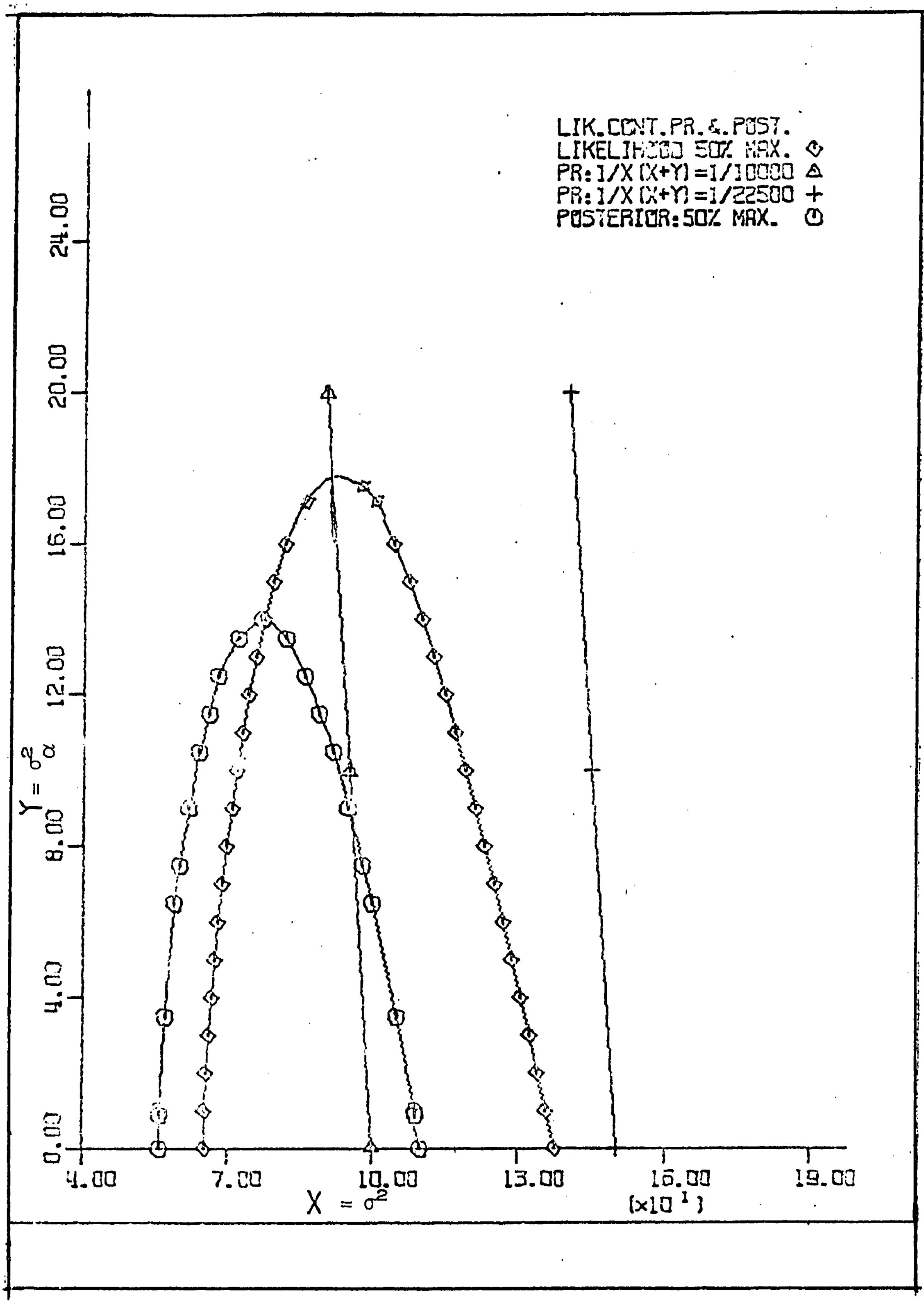

LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX.
PR: 1/X (X+Y) = 1/10000
PR: 1/X (X+Y) = 1/22500
POSTERIOR: 50% MAX.
24.00
20.00
16.00
12.00
8.00
4.00
0.00
$Y = \sigma_\alpha^2$
4.00
7.00
10.00
13.00
16.00
19.00
[x10 1]
$X = \sigma^2$

Figure 17. Contours of 50 percent of maximum for likelihood and posterior with prior $\alpha[\sigma^2(\sigma^2+n\sigma_\alpha^2)]^{-1}$; $k = 4$, $n = 5$; $S_1 = 1,600$, $S_2 = 248$

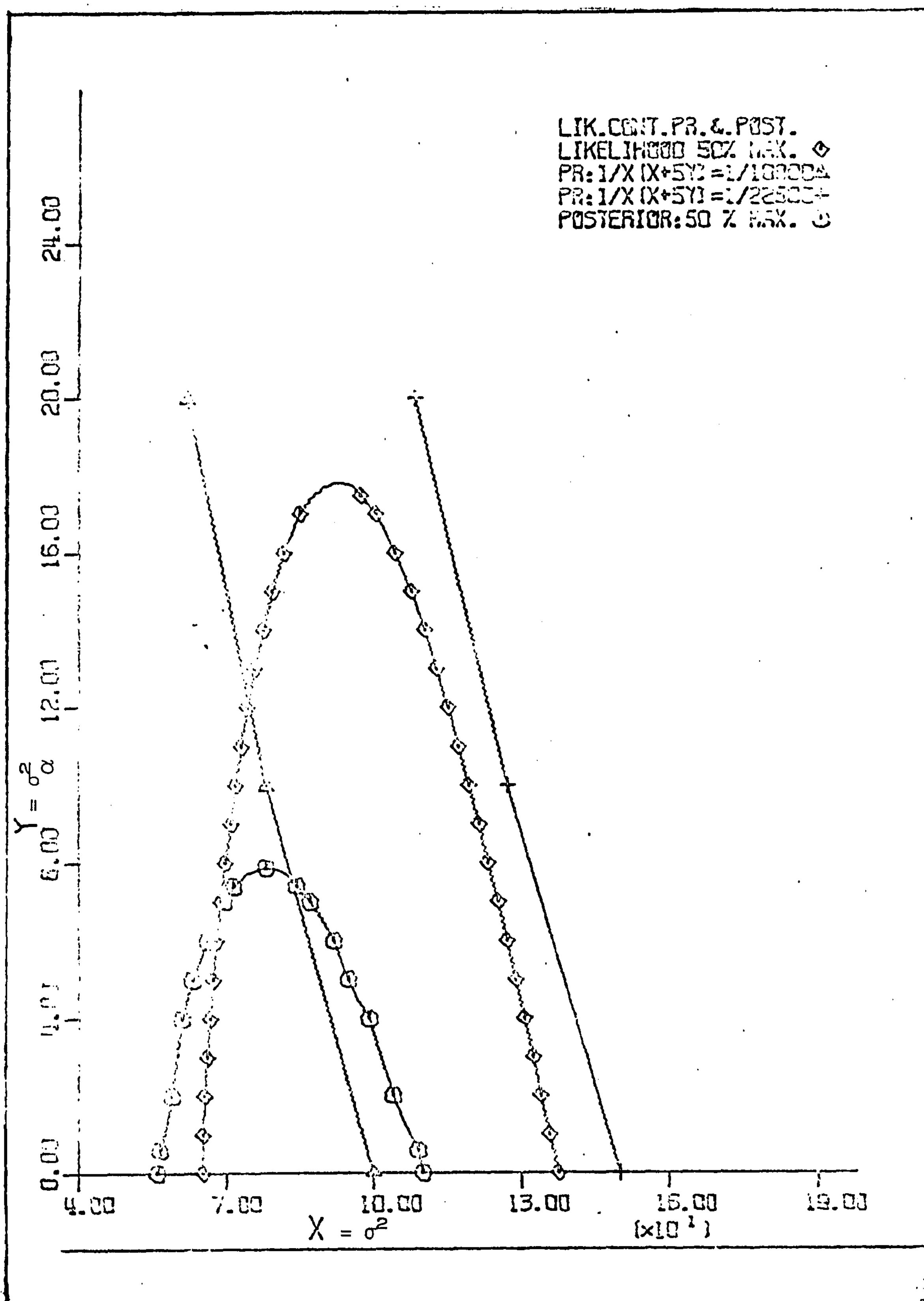

LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX.
PR:1/X (X+5Y)=1/10000
PR:1/X (X+5Y)=1/22500
POSTERIOR:50 % MAX.
Y = σ²α
0.00
4.00
6.00
12.00
16.00
20.00
24.00
X = σ²
4.00
7.00
10.00
13.00
16.00
19.00
(x10 ¹)

Figure 18. Contours of 50 percent of maximum for likelihood and posterior with prior $\alpha(\sigma^2+\sigma_\alpha^2)^{-1}$; $k = 4$, $n = 5$; $S_1 = 1600$, $S_2 = 248$

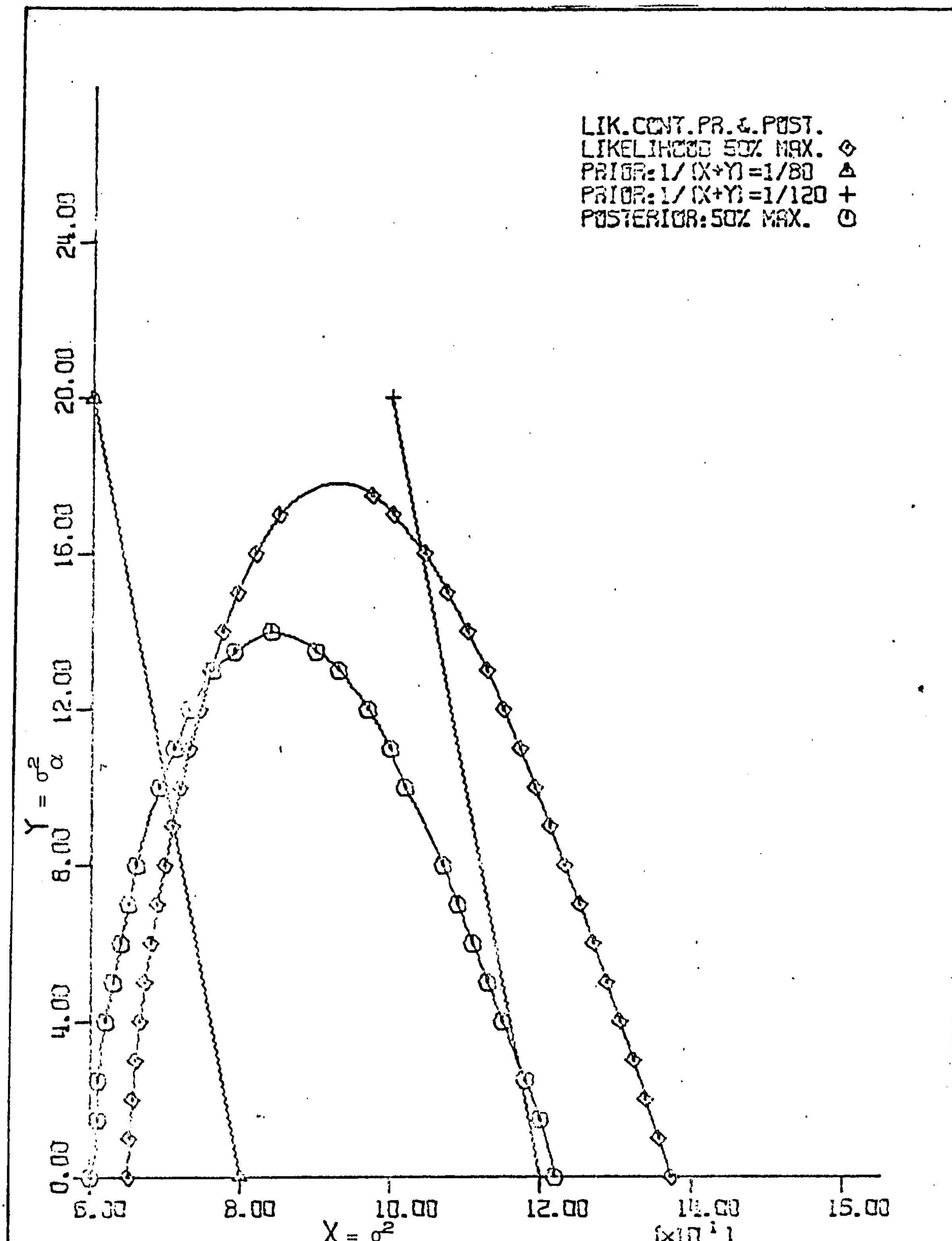

LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX.
PRIOR: 1/(X+Y)=1/80
PRIOR: 1/(X+Y)=1/120
POSTERIOR: 50% MAX.
24.00
20.00
16.00
12.00
8.00
4.00
0.00
$Y = \sigma^2_\alpha$
6.00
8.00
10.00
12.00
14.00
16.00
$X = \sigma^2$
(x10 1)

Figure 19. Contours of 50 percent of maximum for likelihood and posterior with prior $\alpha(\sigma^2+n\sigma_\alpha^2)^{-1}$; $k = 4$, $n = 5$; $S_1 = 1600$, $S_2 = 248$

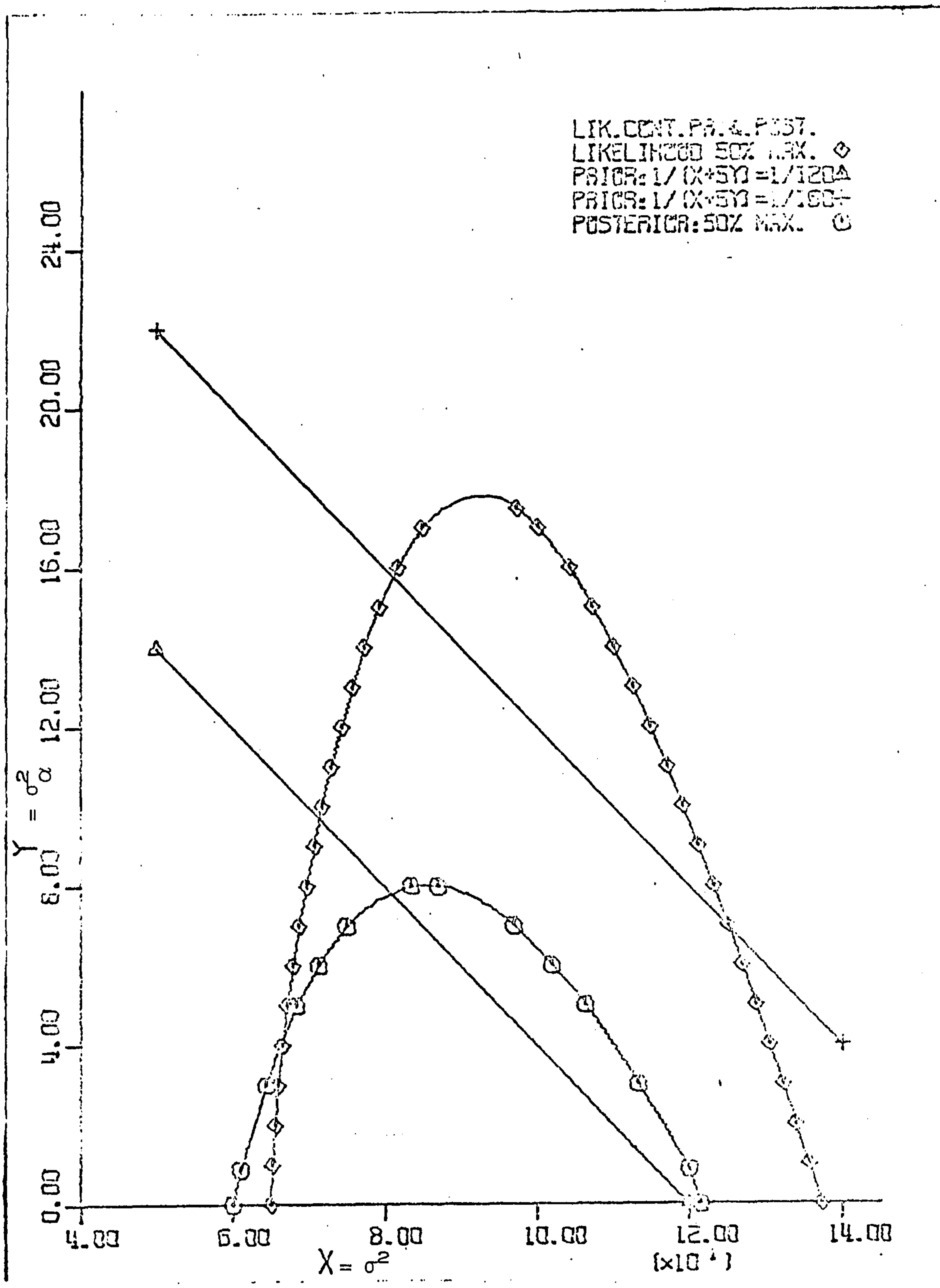
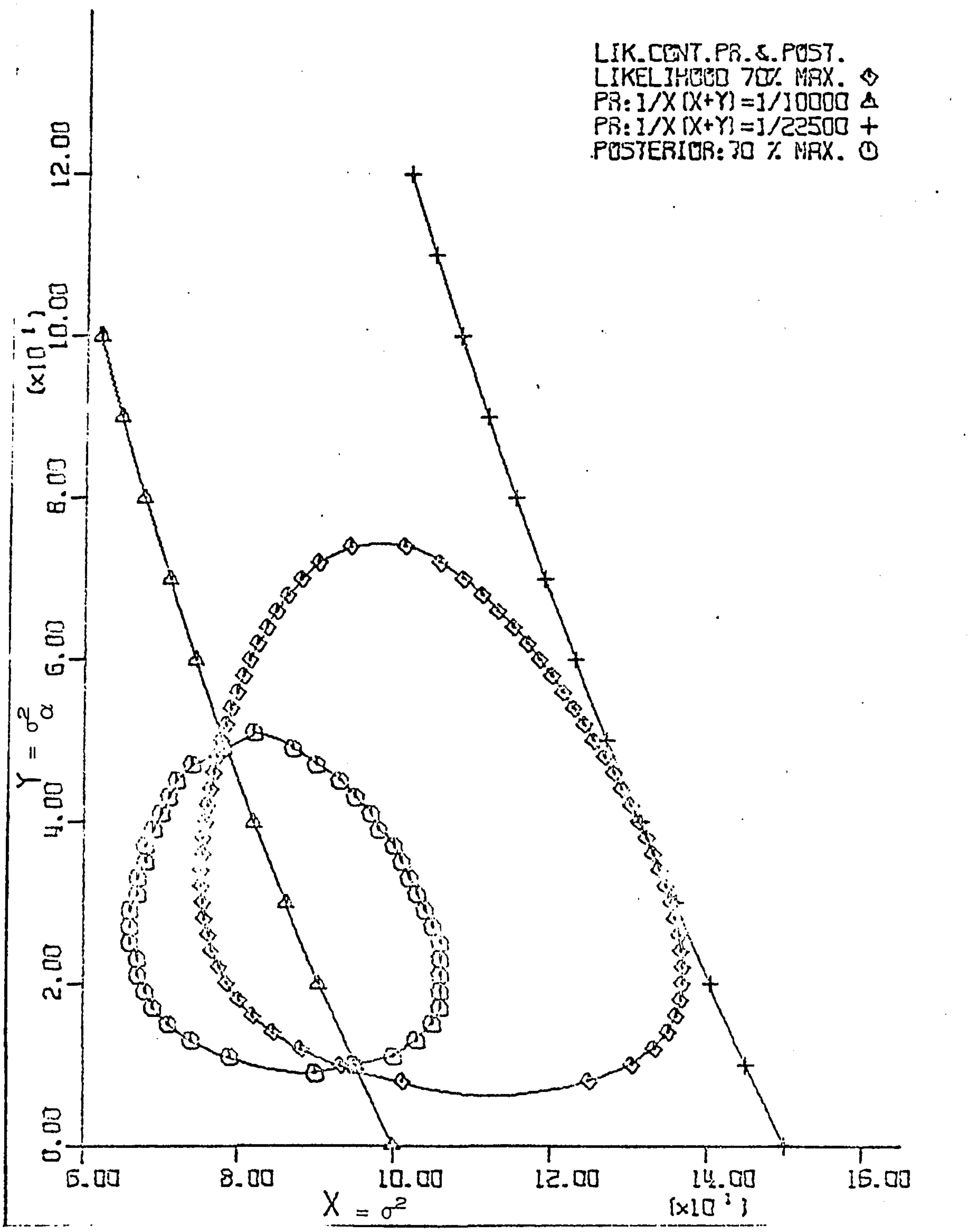
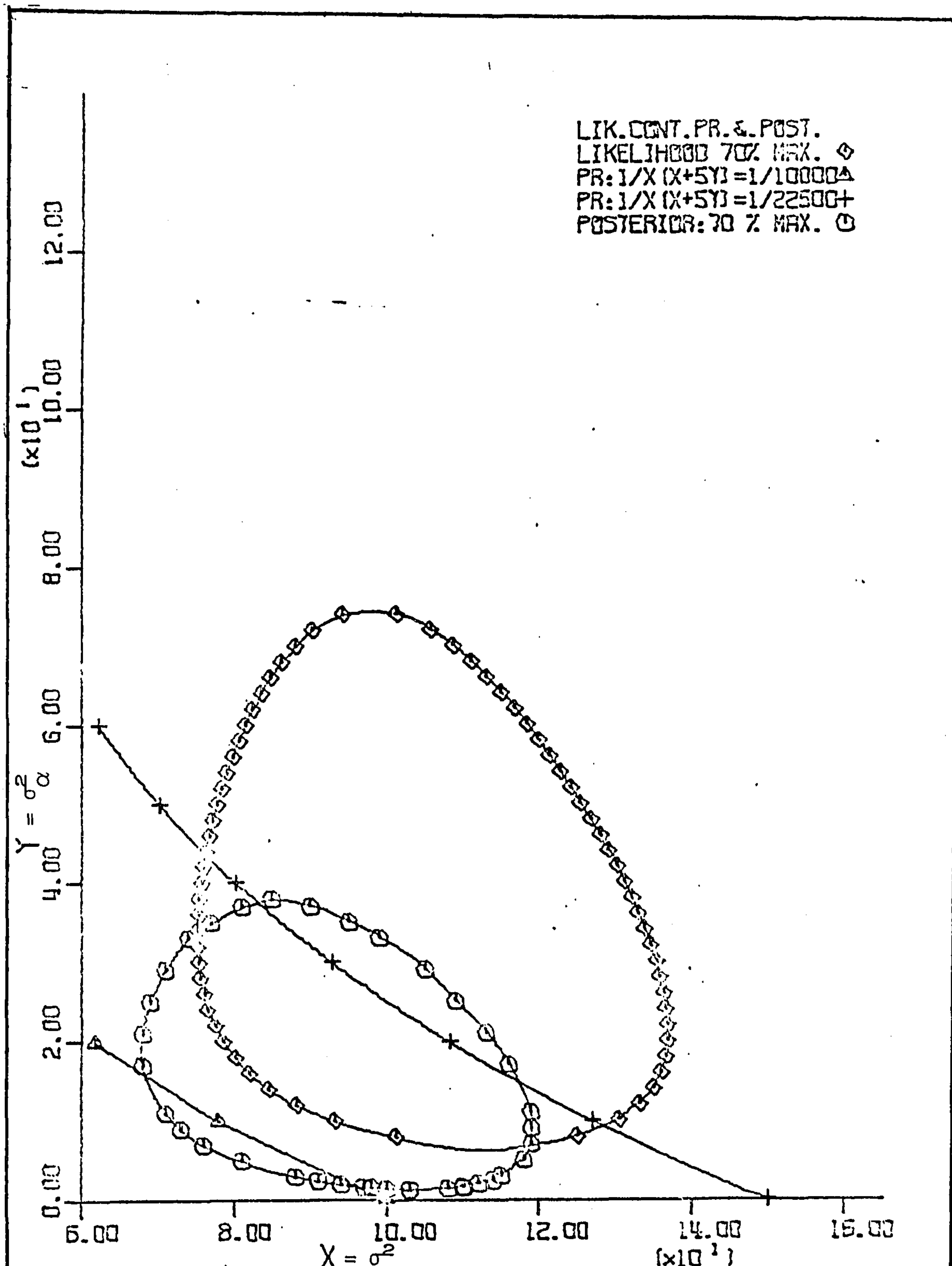

LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX.
PRIOR: 1/(X+5Y)=1/120
PRIOR: 1/(X+5Y)=1/180
POSTERIOR: 50% MAX.
Y = σ_α^2
0.00
4.00
8.00
12.00
16.00
20.00
24.00
X = σ^2
4.00
6.00
8.00
10.00
12.00
14.00
(x10^{1})

Figure 20. Contours of 70 percent of maximum for likelihood and posterior with prior $\alpha[\sigma^2(\sigma^2+\sigma_\alpha^2)]^{-1}$; $k = 4$, $n = 5$; $S_1 = 1600$, $S_2 = 972$



LIK.CONT.PR.&.POST.
LIKELIHOOD 70% MAX.
PR: 1/X (X+Y) =1/10000
PR: 1/X (X+Y) =1/22500
POSTERIOR: 70 % MAX.
Y = σ^2_α
(x10^1)
0.00
2.00
4.00
6.00
8.00
10.00
12.00
X = σ^2
(x10^1)
6.00
8.00
10.00
12.00
14.00
16.00

Figure 21. Contours of 70 percent of maximum for likelihood and posterior with prior $\alpha[\sigma^2(\sigma^2+n\sigma_\alpha^2)]^{-1}$; $k = 4$, $n = 5$; $S_1 = 1600$, $S_2 = 972$



LIK.CONT.PR.&.POST.
LIKELIHOOD 70% MAX.
PR: 1/X (X+5Y) =1/10000
PR: 1/X (X+5Y) =1/22500
POSTERIOR: 70 % MAX.
Y = σ²α
(x10 ¹)
0.00
2.00
4.00
6.00
8.00
10.00
12.00
6.00
8.00
10.00
12.00
14.00
16.00
X = σ²
(x10 ¹)

Figure 22. Contours of 70 percent of maximum for likelihood and posterior with prior $\alpha(\sigma^2+\sigma_\alpha^2)^{-1}$; $k = 4$, $n = 5$; $S_1 = 1600$, $S_2 = 972$

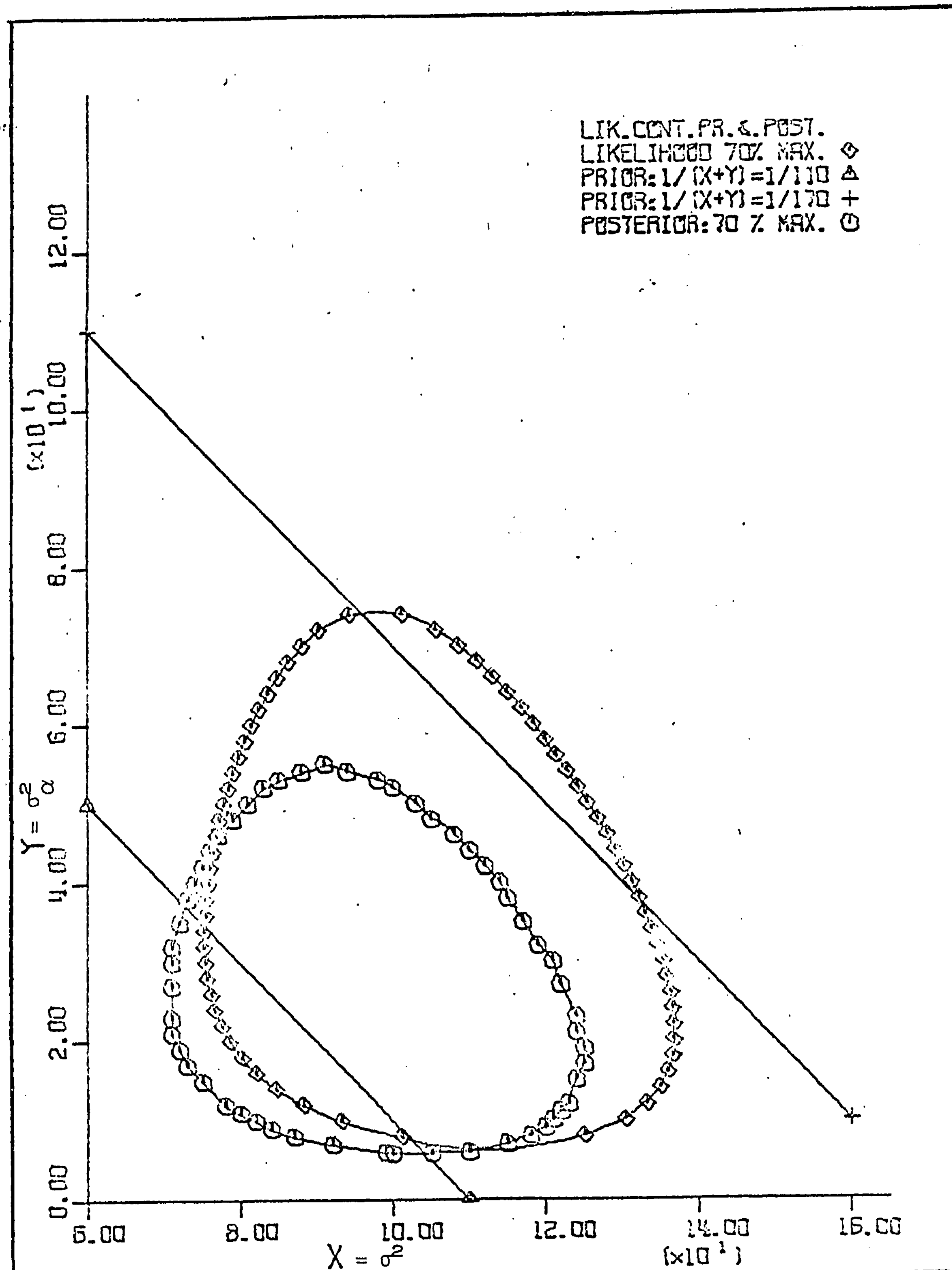

LIK. CONT. PR. & POST.
LIKELIHOOD 70% MAX.
PRIOR: 1/(X+Y)=1/110
PRIOR: 1/(X+Y)=1/170
POSTERIOR: 70 % MAX.
Y = σ_α^2
(x10 1)
0.00
2.00
4.00
6.00
8.00
10.00
12.00
X = σ^2
(x10 1)
6.00
8.00
10.00
12.00
14.00
16.00

Figure 23. Contours of 70 percent of maximum for likelihood and posterior with prior $\alpha(\sigma^2+n\sigma_\alpha^2)^{-1}$; $k = 4$, $n = 5$; $S_1 = 1600$, $S_2 = 972$

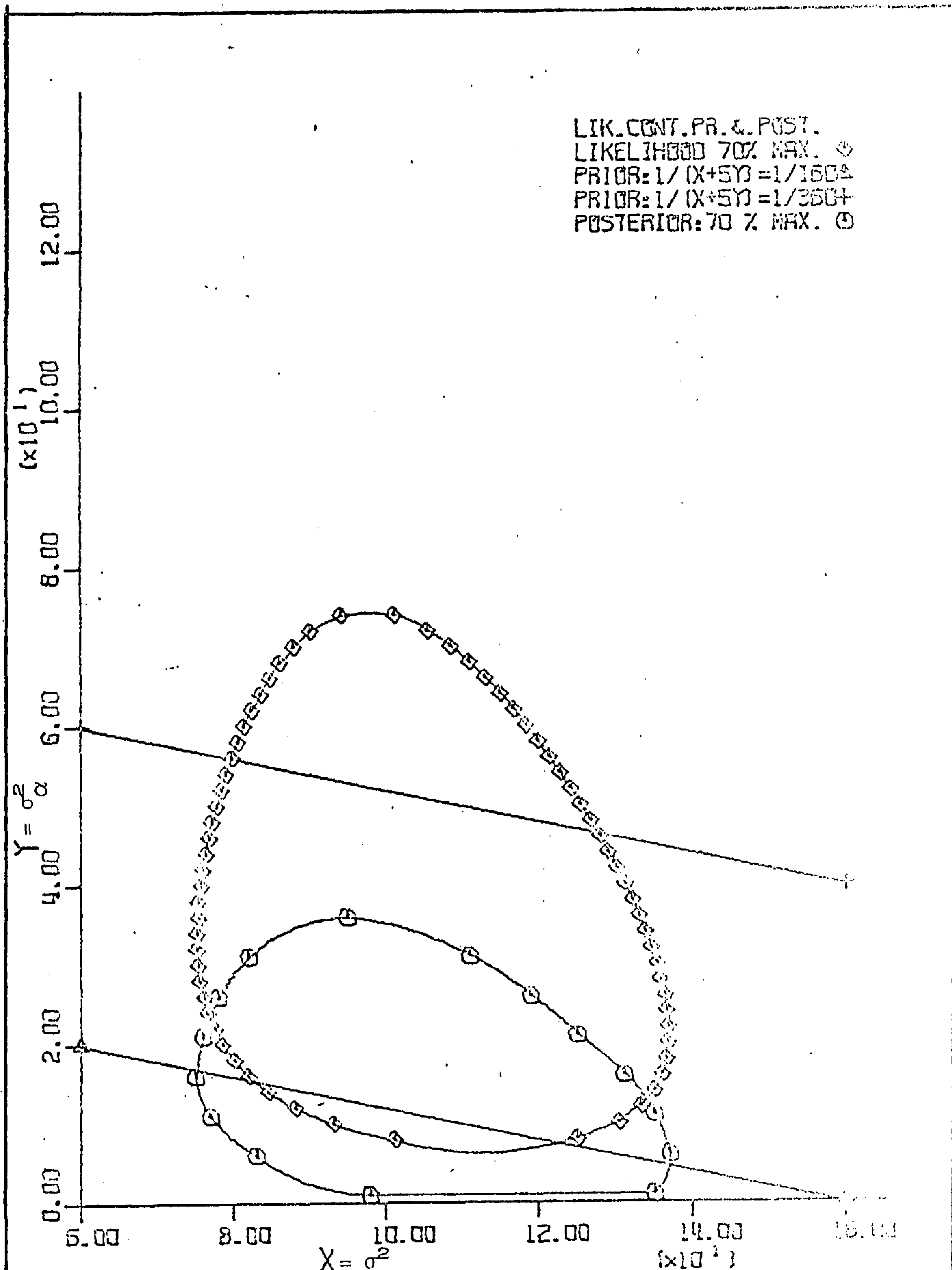

LIK.CONT.PR.&.POST.
LIKELIHOOD 70% MAX.
PRIOR:1/(X+5Y)=1/160
PRIOR:1/(X+5Y)=1/360
POSTERIOR:70 % MAX.
Y = σ_α^2
(x10 1)
0.00
2.00
4.00
6.00
8.00
10.00
12.00
X = σ^2
5.00
8.00
10.00
12.00
14.00
(x10 1)

Figure 24. Contours of 50 percent of maximum likelihood and posterior with prior $\alpha[\sigma^2(\sigma^2+\sigma_\alpha^2)]^{-1}$; $k = 10$, $n = 5$; $S_1 = 8000$, $S_2 = 1850$

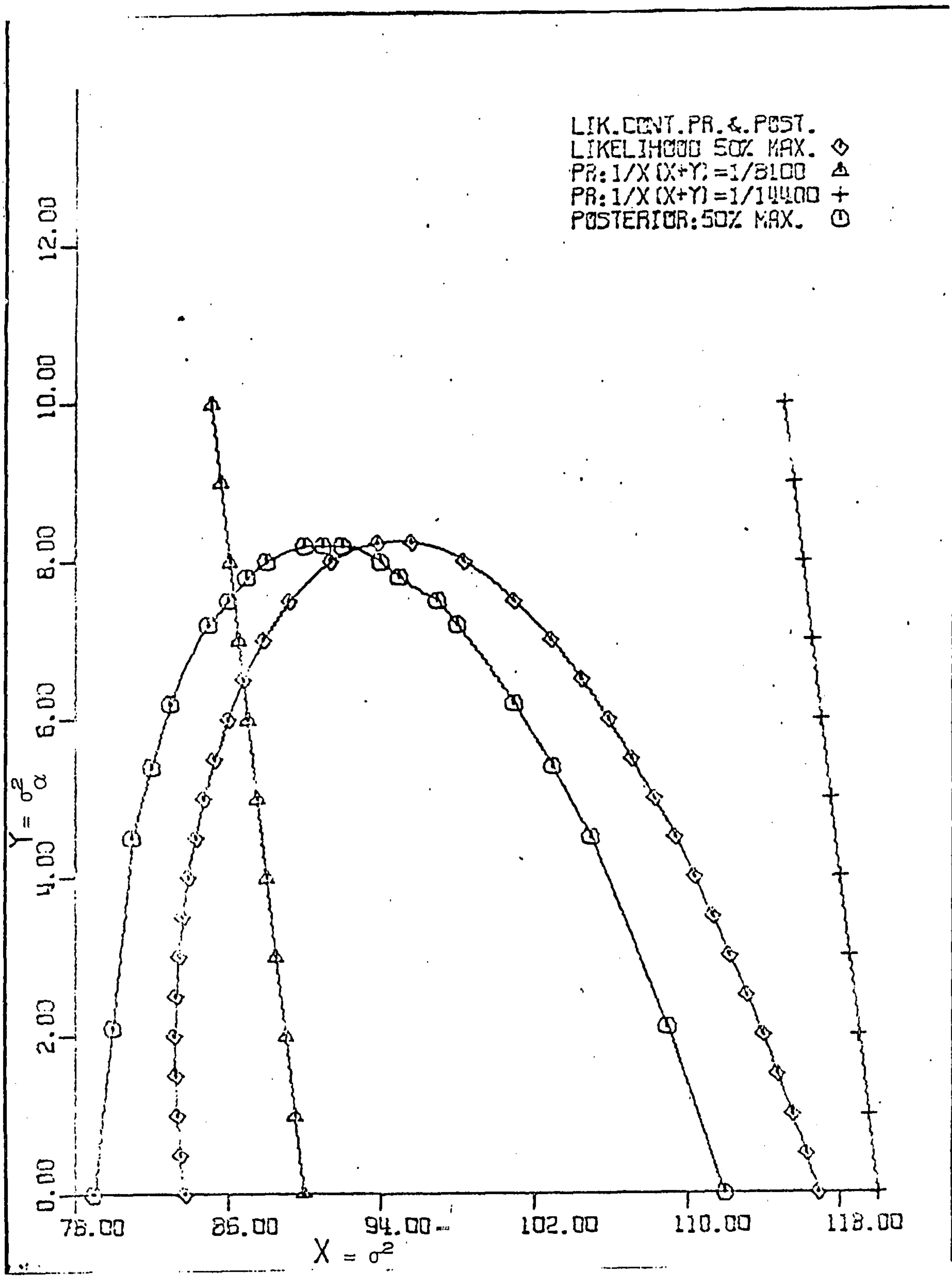
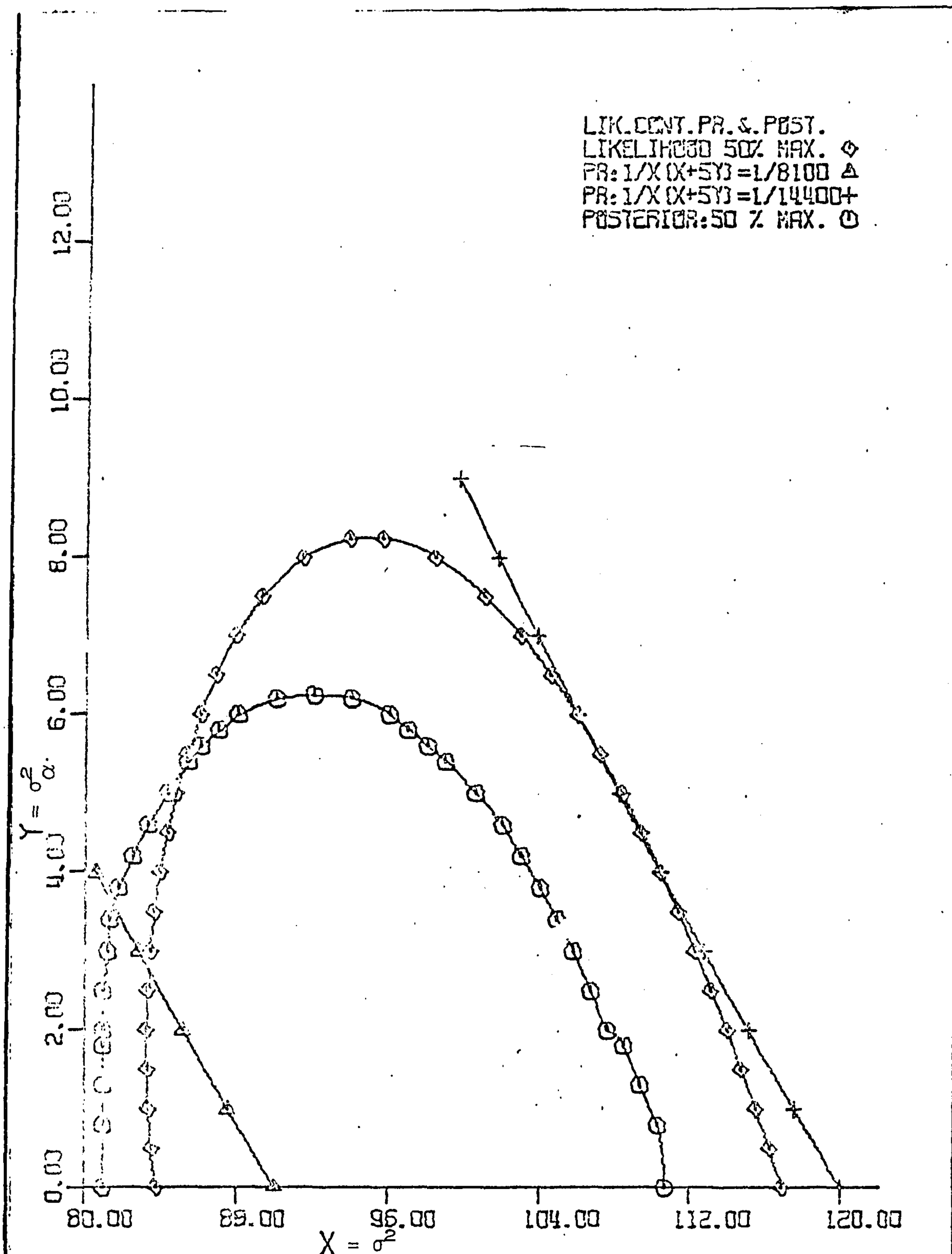

LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX.
PR: 1/X (X+Y) = 1/8100
PR: 1/X (X+Y) = 1/14400
POSTERIOR: 50% MAX.
$Y = \sigma^2_\alpha$
0.00
2.00
4.00
6.00
8.00
10.00
12.00
78.00
86.00
94.00
102.00
110.00
118.00
$X = \sigma^2$

Figure 25. Contours of 50 percent of maximum likelihood and posterior with prior $\alpha[\sigma^2(\sigma^2+n\sigma_\alpha^2)]^{-1}$; $k = 20$, $n = 5$; $S_1 = 8000$, $S_2 = 1850$



LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX.
PR: 1/X (X+5Y) = 1/8100
PR: 1/X (X+5Y) = 1/14400
POSTERIOR: 50 % MAX.
0.00
2.00
4.00
6.00
8.00
10.00
12.00
Y = σ²α
80.00
88.00
96.00
104.00
112.00
120.00
X = σ²

Figure 26. Contours of 50-percent of maximum likelihood and posterior with prior $\alpha(\sigma^2+\sigma_\alpha^2)^{-1}$; k = 20, n = 5; S_1 = 8000, S_2 = 1850

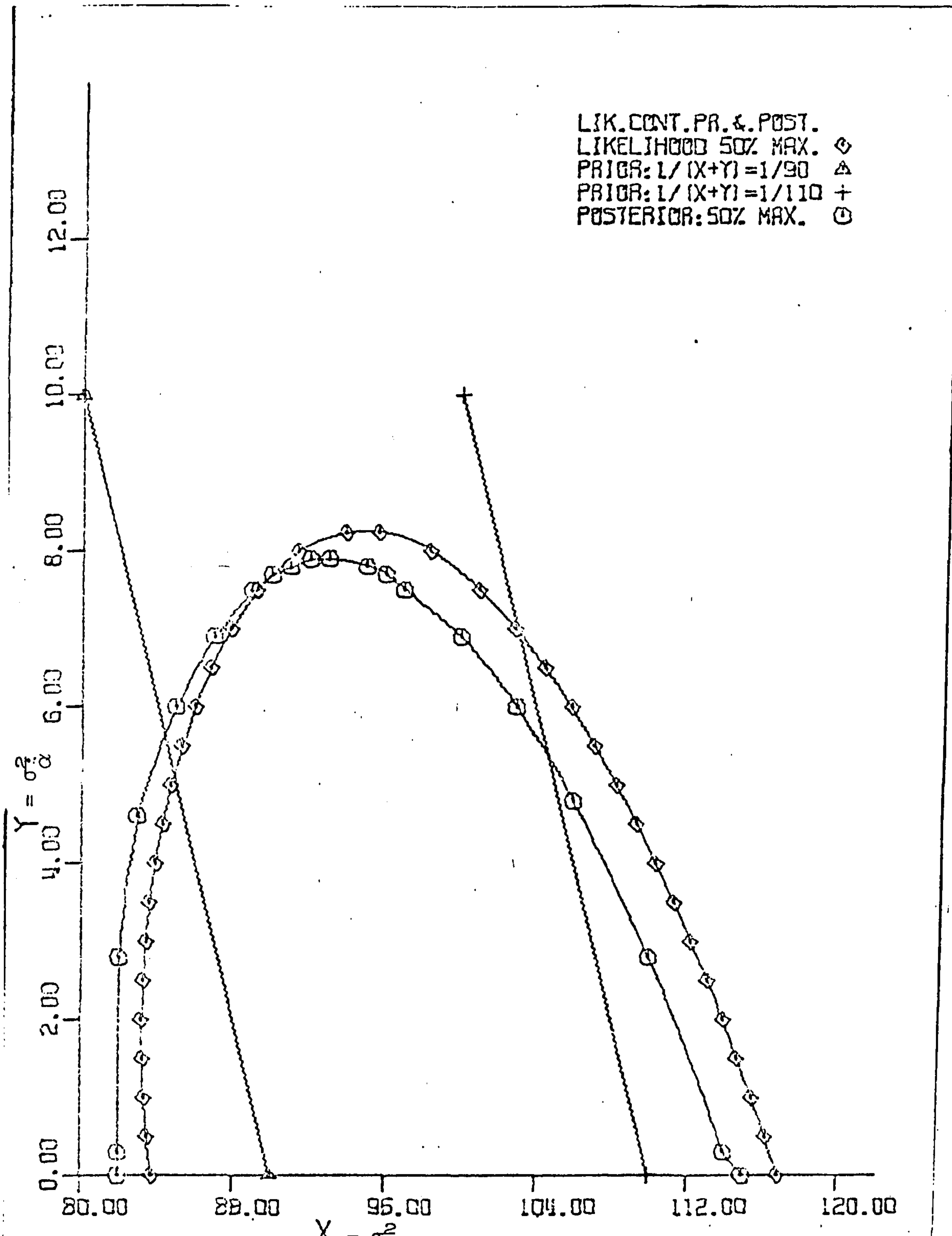

LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX.
PRIOR: 1/(X+Y)=1/90
PRIOR: 1/(X+Y)=1/110
POSTERIOR: 50% MAX.
12.00
10.00
8.00
6.00
4.00
2.00
0.00
$Y = \sigma_\alpha^2$
80.00
88.00
96.00
104.00
112.00
120.00
$X = \sigma^2$

Figure 27. Contours of 50 percent of maximum likelihood and posterior with prior $\alpha(\sigma^2+n\sigma_\alpha^2)^{-1}$; $k = 20$, $n = 5$; $S_1 = 8000$, $S_2 = 1850$

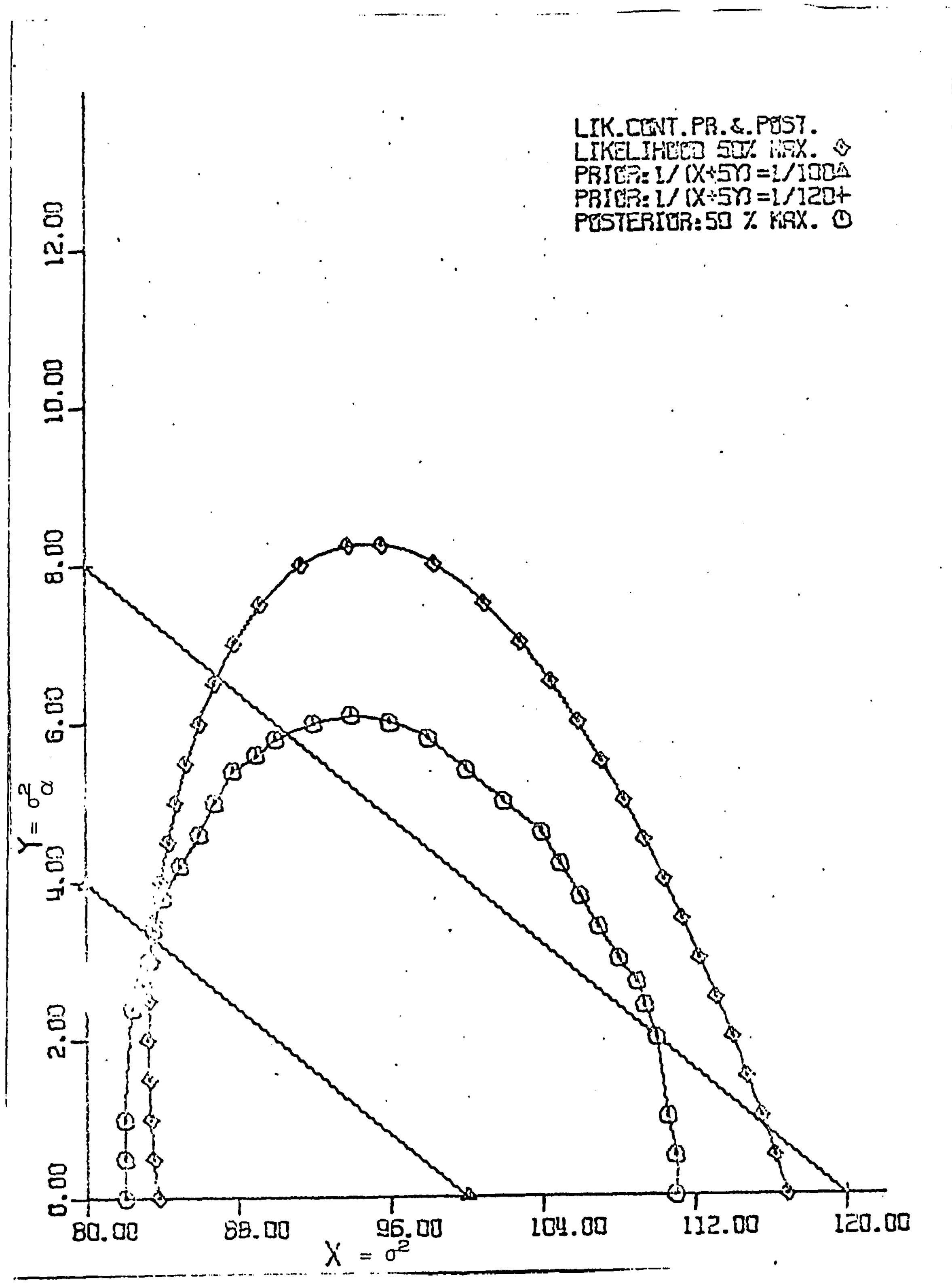

LIK.CONT.PR.&.POST.
LIKELIHOOD 50% MAX. ◇
PRIOR: 1/(X+5Y)=1/100△
PRIOR: 1/(X+5Y)=1/120+
POSTERIOR: 50 % MAX. ⊙
0.00
2.00
4.00
6.00
8.00
10.00
12.00
Y = σ_α^2
80.00
88.00
96.00
104.00
112.00
120.00
X = σ^2

Figure 28. Contours of 70 percent of maximum likelihood and posterior with prior $\alpha[\sigma^2(\sigma^2+\sigma_\alpha^2)]^{-1}$; $k = 20$, $n = 5$; $S_1 = 8000$, $S_2 = 3268$

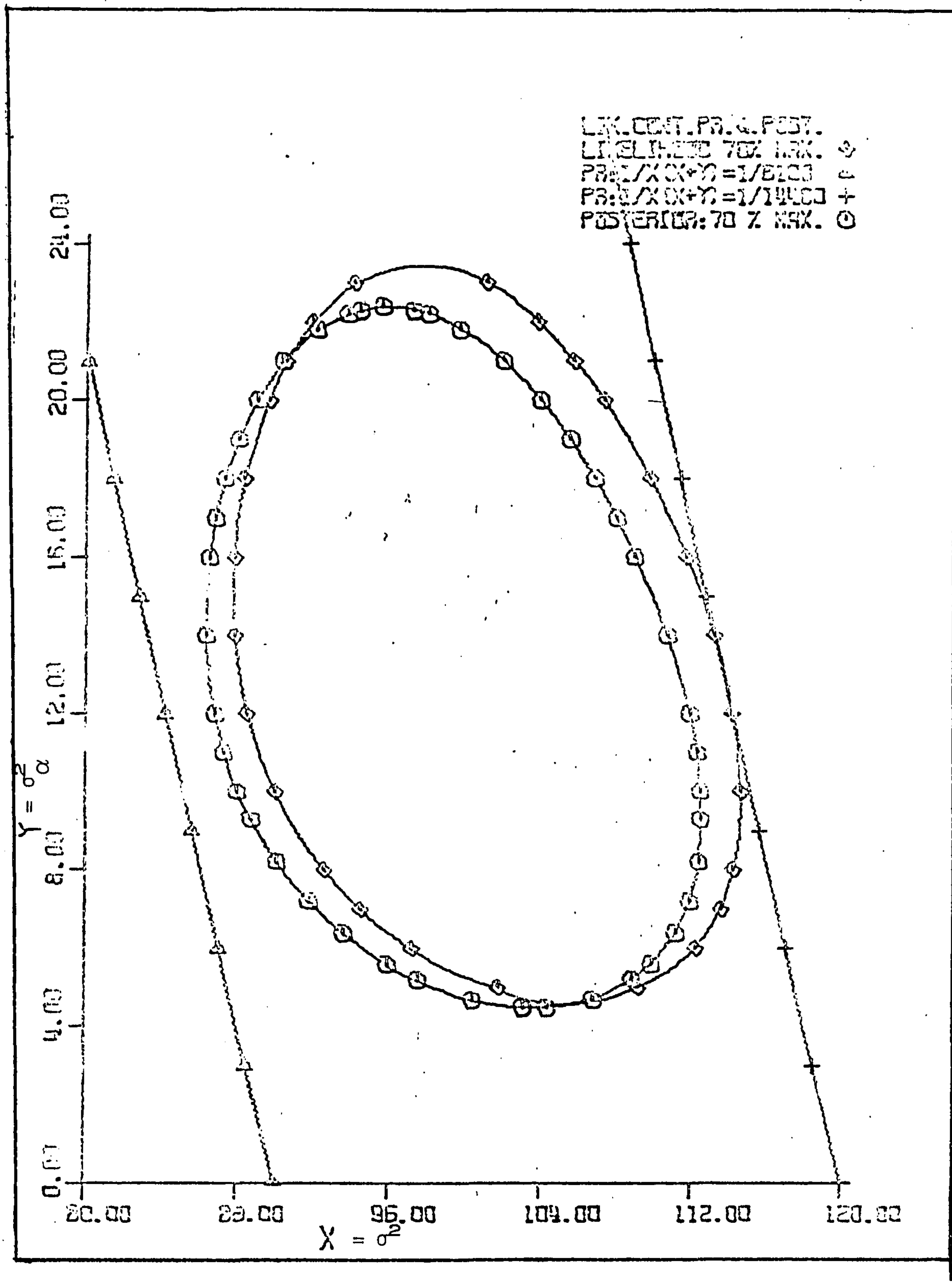

LIK.CONT.PR.&.POST.
LIKELIHOOD 70% MAX.
POSTERIOR: 70 % MAX.
24.00
20.00
16.00
12.00
8.00
4.00
0.00
Y = σ²α
80.00
88.00
96.00
104.00
112.00
120.00
X = σ²

Figure 29. Contours of 70 percent of maximum likelihood and posterior with prior $\alpha[\sigma^2\epsilon\sigma^2+n\sigma^2_\alpha)]^{-1}$; $k = 20$, $n = 5$; $S_1 = 8000$, $S_2 = 3268$

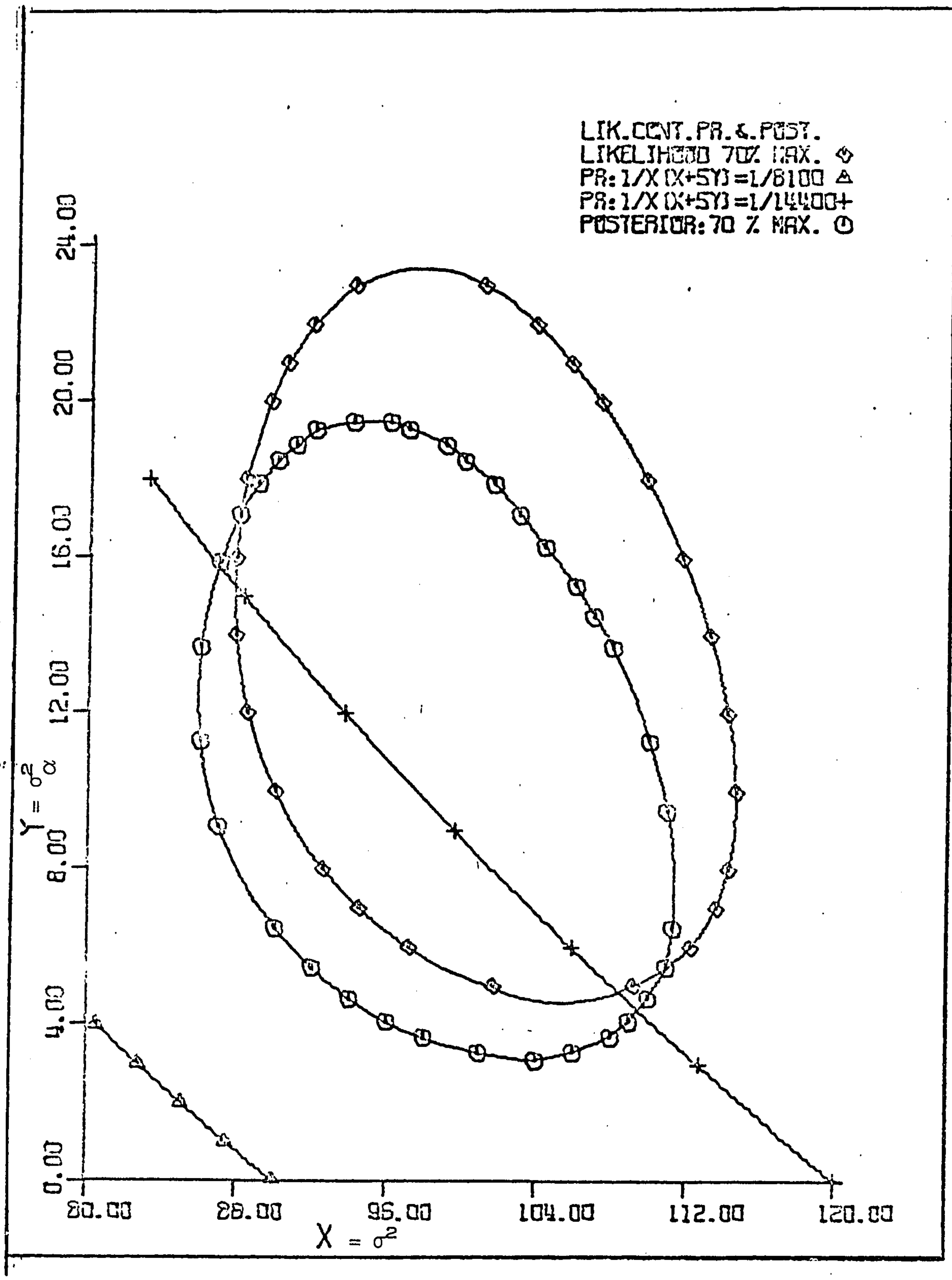

LIK.CCNT.PR.&.POST.
LIKELIHOOD 70% MAX.
PR: 1/X (X+5Y) = 1/8100
PR: 1/X (X+5Y) = 1/14400+
POSTERIOR: 70 % MAX.
24.00
20.00
16.00
12.00
8.00
4.00
0.00
$Y = \sigma^2_\alpha$
80.00
88.00
96.00
104.00
112.00
120.00
$X = \sigma^2$

Figure 30. Contours of 70 percent of maximum likelihood and posterior with prior $\alpha(\sigma^2+\sigma_\alpha^2)^{-1}$; $k = 20$, $n = 5$; $S_1 = 8000$, $S_2 = 3268$

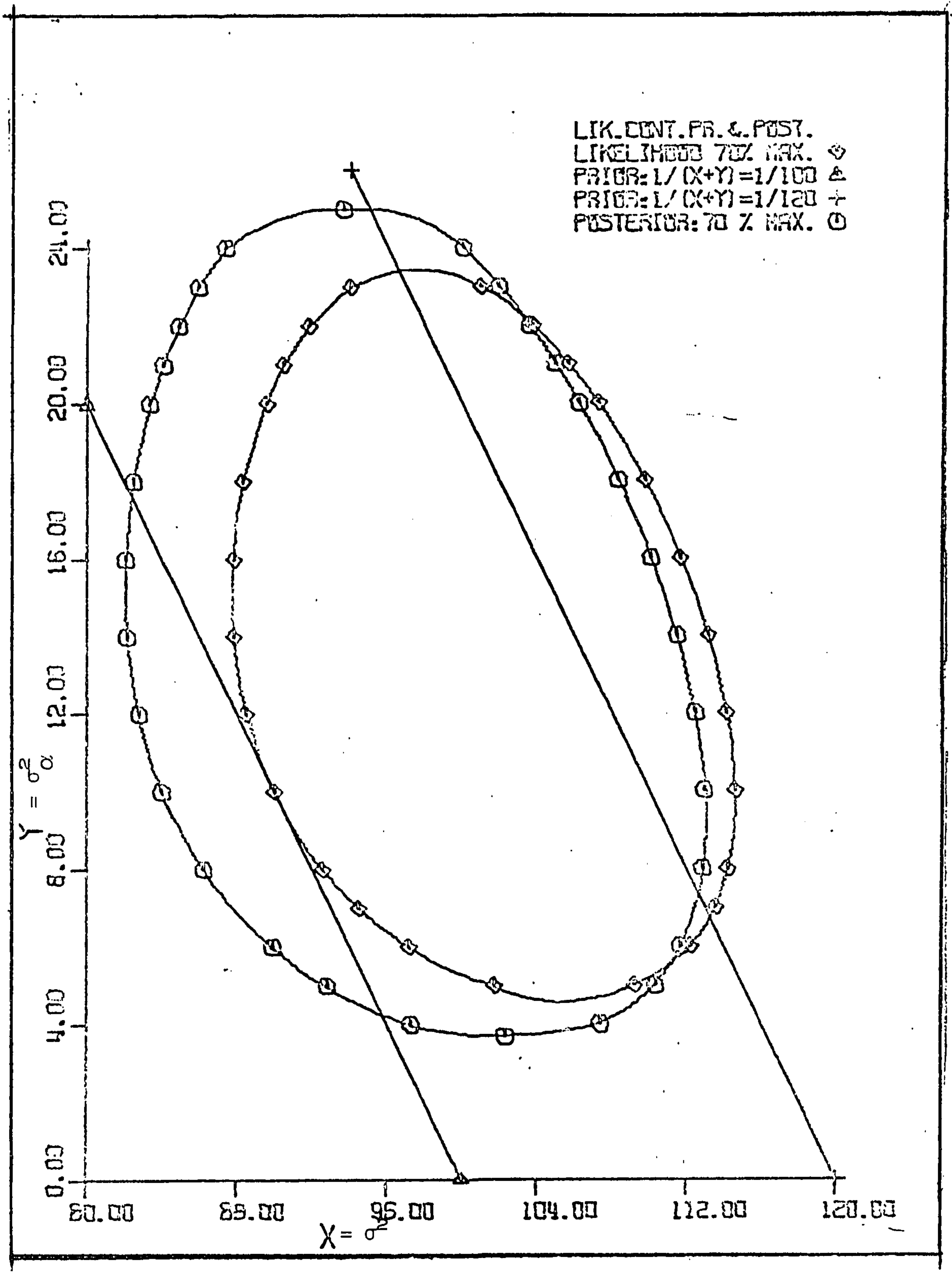
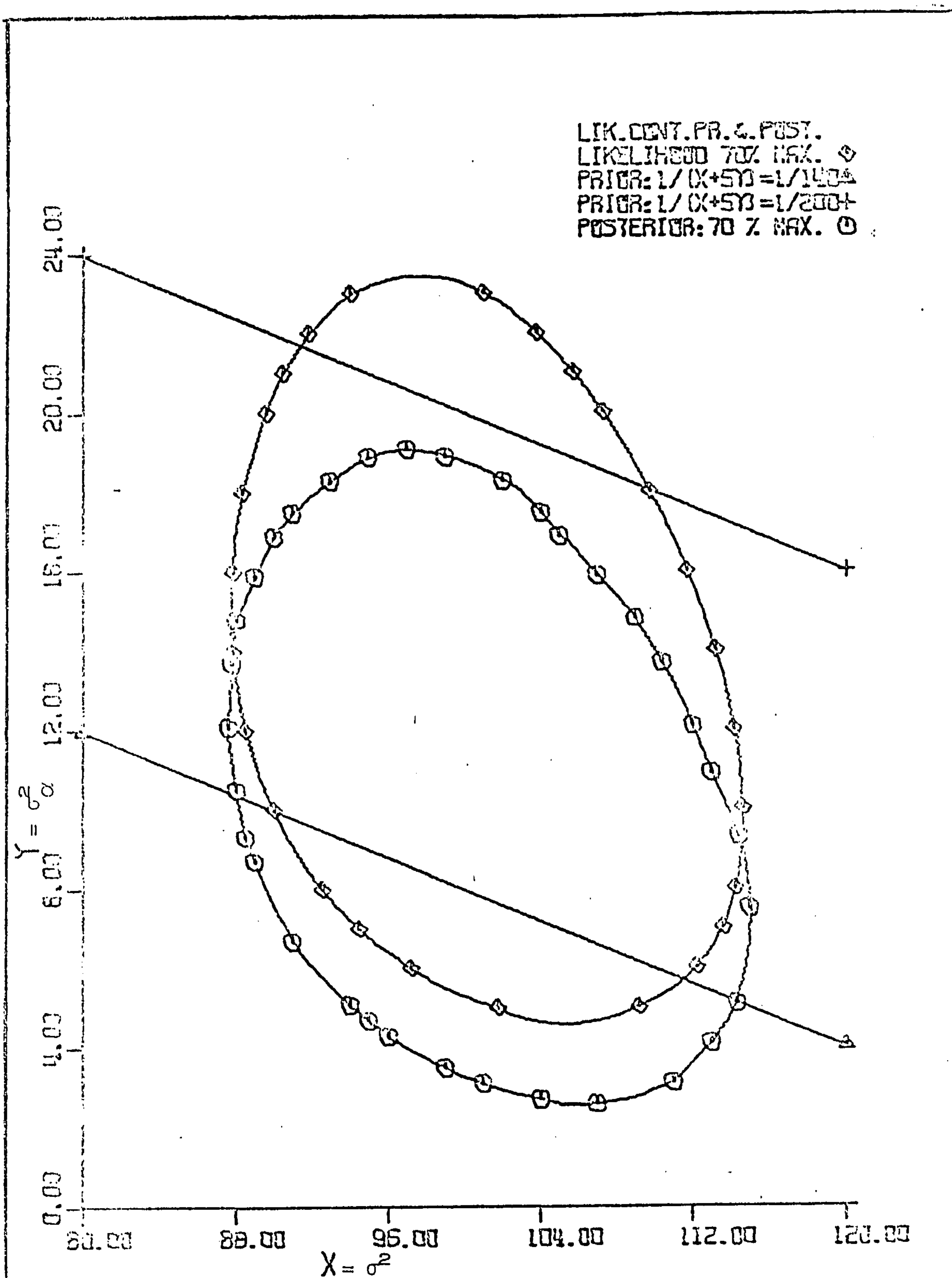

LIK.CONT.PR.&.POST.
LIKELIHOOD 70% MAX.
PRIOR: 1/(X+Y)=1/100
PRIOR: 1/(X+Y)=1/120
POSTERIOR: 70 % MAX.
24.00
20.00
16.00
12.00
8.00
4.00
0.00
80.00
88.00
96.00
104.00
112.00
120.00
Y = σ²α
X = σ²

Figure 31. Contour of 70 percent of maximum likelihood and posterior with prior $\alpha(\sigma^2+n\sigma_\alpha^2)^{-1}$; $k = 20$, $n = 5$; $S_1 = 8000$, $S_2 = 3268$


LIK.CONT.PR.&.POST.
LIKELIHOOD 70% MAX.
PRIOR: 1/(X+5Y)=1/140
PRIOR: 1/(X+5Y)=1/200
POSTERIOR: 70 % MAX.
24.00
20.00
16.00
12.00
4.00
0.00
Y = σ²α
80.00
88.00
96.00
104.00
112.00
120.00
X = σ²

V. GOODNESS OF FIT

The model considered by us in the present study is

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad (i=1, \ldots k,\ j=1, \ldots n)$$

$\alpha_i \sim NID(0,\sigma_\alpha^2)$; $\epsilon_{ij} \sim NID(0,\sigma^2)$; α_i and ϵ_{ij} independent.

Denoting

$$\bar{y}_{i.} = \sum_{j=1}^{n} y_{ij}/n \quad ; \quad \bar{y}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}/kn$$

$$S_1 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij}-\bar{y}_{i.})^2 \quad ; \quad S_2 = n\sum_{i=1}^{k} (\bar{y}_{i.}-y_{..})^2 \quad .$$

We have observed that the likelihood function, after maximizing for μ, is

$$\underset{\text{Max}/\mu}{L}(\sigma^2,\sigma_\alpha^2/Y)\alpha[(\sigma^2)^{-\frac{k(n-1)}{2}} \text{Exp}(-\frac{S_1}{2\sigma^2})][(\sigma^2+n\sigma_\alpha^2)^{-\frac{k}{2}} \text{Exp}(-\frac{S_2}{2(\sigma^2+n\sigma_\alpha^2)})] \ .$$

It is evident that the likelihood function depends on σ^2 and $(\sigma^2+n\sigma_\alpha^2)$ and it can not be written as a product of a function of σ^2 alone and a function of σ_α^2 alone. The likelihood function does not factorize i.e., there exist no functions of the form $M(\sigma^2)$ and $N(\sigma_\alpha^2)$ such that

$$\underset{\text{Max}/\mu}{L}(\sigma^2,\sigma_\alpha^2) = M(\sigma^2)N(\sigma_\alpha^2) \quad .$$

We agree with Barnard et al. (2) that when the likelihood function does not factorize, separate consideration of parameters may not be possible. This fact has to be faced and any attempt to ignore it should be resisted. If, for instance, there are two parameters θ_1 and θ_2 which are not functional related, and with possible values lying in a product space, the separability can be effected. But in the case of two parameters $\theta_1 = \sigma^2$ and $\theta_2 = \sigma^2 + n\sigma_\alpha^2$, it is clear that $\theta_2 \geq \theta_1$, and the space of θ_2 dependes on θ_1. It seems clear that separate inference on σ^2 and $\sigma^2 + n\sigma_\alpha^2$ can not be made. If attention is directed to "point estimation", it is reasonable to require that the estimates should be functionally consistent. It therefore seems that ideas of point estimation can be applied only with the use of a loss function which involves $\hat{\sigma}^2$, $\hat{\sigma}_\alpha^2$, σ^2 and σ_α^2. Unbiassedness or minimum mean square error of estimation of one parameter, such as σ_α^2, seems to be irrelevant except under special terminal decision situations.

Barnard et al. (2) have attempted to obtain the likelihood of one parameter, which they call pseudo-likelihood, through the process of elimination. Let $L(\alpha,\beta)$ be an unfactorizable likelihood involving two parameters α,β. The pseudo-likelihood $M(\alpha)$ involving only α, eliminating β is

$$M(\alpha) = E\{L(\alpha,\beta)\} = \int L(\alpha,\beta)dW(\beta)$$

where

(i) $E\{\ \}$ denotes the operation of eliminating β which, under certain conditions imposed by them, is a linear functional and

(ii) $W(\beta)$ is some measure function defined over the values of β.

They remark that the only way of eliminating parameters we do not want to discuss is to integrate them out. It is quite clear that the whole procedure is essentially a Bayesian approach with $W(\beta)$ as the prior distribution of β. In the present study, the parameters σ^2 and σ_α^2 are not separable. We feel, strongly therefore, that if we do not want to be Bayesian, we should attempt a joint inference about $(\sigma^2,\sigma_\alpha^2)$. In this chapter we propose and explain a measure of goodness of fit of observed data with respect to a given pair $(\sigma^2,\sigma_\alpha^2)$.

Suppose that an experimenter has performed an experiment with the assumed model as above and obtains the sample values (S_{10},S_{20}). We ask ourselves the question that given the observed data, condensed in (S_{10},S_{20}), what is the goodness of fit of the data with respect to a given pair of parameters $(\sigma^2,\sigma_\alpha^2)$. A general idea of testing goodness of fit, which goes back a long time in the history of statistical ideas and was the explicit basis of the Pearson χ^2 method, is to order the possible sample results on the basis of their probabilities. This has been used many times by Fisher in references too numerous to quote and by Neyman and Pearson (27, 28). The goodness of fit is defined as

$$\text{G.F.}(S_{10},S_{20};\sigma^2,\sigma_\alpha^2) = \iint P(S_1,S_2/\sigma^2,\sigma_\alpha^2)\ \psi\ (S_1,S_2;S_{10},S_{20};\sigma^2,\sigma_\alpha^2)dS_1,dS_2$$

where

$$\begin{aligned}\psi(S_1,S_2;S_{10},S_{20};\sigma^2,\sigma_\alpha^2) &= 1\ ,\ \text{if}\quad P(S_1,S_2;\sigma^2,\sigma_\alpha^2) \le P(S_{10},S_{20};\sigma^2,\sigma_\alpha^2)\\ &= 0 \quad \text{otherwise.}\end{aligned}$$

According to this measure, we can calculate a goodness of fit value of (S_{10},S_{20}) with respect to any pair $(\sigma^2,\sigma_\alpha^2)$. In $(\sigma^2,\sigma_\alpha^2)$ plane, we can plot this against $(\sigma^2,\sigma_\alpha^2)$ and thus obtain contours of equal goodness of fit value. Similarly, for a fixed pair $(\sigma^2,\sigma_\alpha^2)$, we can calculate a goodness of fit value of any (S_1,S_2) pair and plot this against (S_1,S_2) in

(S_1,S_2) plane. We can, thus, have a system of contours of equal goodness of fit value with center at $S_1 = \sigma^2[k(n-1)-2]$ and $S_2 = (k-3)(\sigma^2+n\sigma_\alpha^2)$. It is easy to verify that all pairs (S_1,S_2) will have the same goodness of fit with respect to $(\sigma^2,\sigma_\alpha^2)$ on the following contour

$$[k(n-1)-2] \log S_1 + (k-3) \log S_2 - \frac{S_1}{\sigma^2} - \frac{S_2}{(\sigma^2+n\sigma_\alpha^2)} = C_o$$

where C_o is an admissible constant.

In order that the above principle of goodness of fit may be applicable, in general, we propose to work in a different co-ordinate system.

Let

$$W = \frac{S_1}{\sigma^2} \; ; \; B = \frac{S_2}{(\sigma^2+n\sigma_\alpha^2)}$$

then

$$W \sim \chi^2_{k(n-1)} \quad \text{and} \quad B \sim \chi^2_{(k-1)}$$

and the joint density function is

$$f(W,B) = [\Gamma(\tfrac{k-1}{2})\,\Gamma(\tfrac{k(n-1)}{2})\, 2^{(kn-1)/2}]^{-1} W^{\frac{k(n-1)-2}{2}} B^{\frac{k-3}{2}} \times$$

$$\text{Exp}[-\tfrac{1}{2}(W+B)]$$

$$W \geq 0 \; , \; B \geq 0 \; .$$

Suppose that we have observed the data (S_{10},S_{20}) and wish to find the goodness of fit value with respect to $(\sigma_o^2,\sigma_{\alpha o}^2)$. Now $(S_{10},S_{20};\sigma_o^2,\sigma_{\alpha o}^2)$ uniquely determine $W_o = \frac{S_{10}}{\sigma_o^2}$, $B_o = \frac{S_{20}}{(\sigma_o^2+n\sigma_{\alpha o}^2)}$. We shall, then, call the pair (W_o,B_o) as the observed value of (W,B). The goodness of fit value associated with (W_o,B_o) will, then, be

$$\iint f(W,B)\ \psi\ (W,B;W_o,B_o)dWdB$$

where

$$\psi(W,B;W_o,B_o) = 1 \quad \text{if} \quad f(W,B) \le f(W_o,B_o)$$

$$= 0 \quad \text{otherwise.}$$

This can be expressed in the following form

$$G.F(W_o,B_o) =$$

$$\left[\Gamma\left(\frac{(k-1)}{2}\right)\Gamma\left(\frac{k(n-1)}{2}\right)2^{(kn-1)/2}\right]^{-1}\int_{\omega} W^{\frac{k(n-1)-2}{2}} B^{\frac{k-3}{2}} \operatorname{Exp}[-\tfrac{1}{2}(W+B)]dWdB$$

where ω is the region of integration defined by

$$W^{\frac{k(n-1)-2}{2}} B^{\frac{k-3}{2}} \operatorname{Exp}[-\tfrac{1}{2}(W+B)] \le W_o^{\frac{k(n-1)-2}{2}} B_o^{\frac{k-3}{2}} \operatorname{Exp}[-\tfrac{1}{2}(W_o+B_o)]$$

or

$$[k(n-1)-2]\log W + (k-3)\log B - (W+B) \le [k(n-1)-2]\log W_o + (k-3)\log B_o - (W_o+B_o)$$

$$= C_o \quad \text{say.}$$

It is clear that all pairs (W,B), satisfying the following equation, will have the same goodness of fit as (W_o,B_o).

$$[k(n-1)-2]\log W + (k-3)\log B - (W+B) = C_o$$

where C_o is an admissible constant.

For different values of C_o, we shall have contours of equi-goodness of fit in $(W\text{-}B)$ plane. The system of contours will have center at $W = k(n-1)-2$, $B = k-3$ and C_o will have its maximum value,

$[k(n-1)-2]\log[k(n-1)-2] + (k\ 3)\log(k-3) - (kn-5)$, at this point. The advantage of working in (W-B) is now clear. We have reduced the four-dimensional space $(S_1,S_2;\sigma^2,\sigma^2_\alpha)$ into the two dimensional space (W,B) such that for every point $(S_{10},S_{20};\sigma^2_o,\sigma^2_{\alpha o})$ we have a unique point (W_o,B_o). A table of goodness of fit value at any desired level α has only two entries, viz. n and k. Given (n,k) and the desired level of goodness of fit α, we can read the corresponding value C_α from a table. If the given $(S_{10},S_{20};\sigma^2_o,\sigma^2_{\alpha o})$ determines (W_o,B_o) then

$$\text{G.F.}(S^2_{10},S^2_{20};\sigma^2_o,\sigma^2_{\alpha o}) \lesseqgtr \alpha \text{ if } [k(n-1)-2]\log W_o + (k-3)\log B_o - (W_o+B_o) \lesseqgtr C_\alpha .$$

If exact goodness of fit value is required then numerical analysis methods are available to evaluate the necessary integral.

Asymptotic Case I: k large If k is large enough to justify normal approximations of $\chi^2_{(k-1)}$ and $\chi^2_{k(n-1)}$, then the evaluation of goodness of fit value of the observed $(S_{10};S_{20})$ with respect to a given $(\sigma^2_o,\sigma^2_{\alpha o})$ becomes very simple and straight-forward. Using Fisher's approximation, we have

$$\sqrt{\frac{2S_1}{\sigma^2}} = \sqrt{2W} \simeq N(\sqrt{[2k(n-1)-1]},1)$$

and

$$\sqrt{\frac{2S_2}{(\sigma^2+n\sigma^2_\alpha)}} = \sqrt{2B} \simeq N(\sqrt{(2k-3)},1) .$$

Let

$$W' = \sqrt{2W} - \sqrt{[2k(n-1)-1]}$$

$$B' = \sqrt{2B} - \sqrt{(2k-3)} .$$

Then W' and B' are independent standard normal variates. Given $(S_{10},S_{20};\sigma_o^2\sigma_{\alpha o}^2)$, we can uniquely determine a pair (W_o',B_o') in the W'-B' plane. Working in this coordinate system, we have

$$G.F.(W_o',B_o') = \frac{1}{2\pi}\int_\omega \mathrm{Exp}[-\tfrac{1}{2}(W'^2+B'^2)]dW'dB'$$

where ω is the region of integration defined by

$$W'^2+B'^2 \geq W_o'^2 + B_o'^2 =$$

$$[\sqrt{(\frac{2S_{10}}{\sigma_o^2})} - \sqrt{[2k(n-1)-1]}\ ]^2 + [\sqrt{(\frac{2S_{20}}{\sigma_o^2+n\sigma_{\alpha o}^2})} - \sqrt{(2k-3)}]^2$$

$$= C_o'^2 \text{ , say.}$$

In (W',B') plane, the contours of equi-goodness of fit are circles with the center at the origin, and the minimum value of $C_o^{2\prime}$ is zero. It is easy to verify that

$$G.F.(W_o',B_o') = \int\int_\omega \mathrm{Exp}[-\tfrac{1}{2}(W'^2+B'^2)]dW'dB'$$

$$= \mathrm{Exp}[-\tfrac{1}{2}(W_o'^2+B_o'^2)] = \mathrm{Exp}(-\tfrac{1}{2}C_o'^2) \quad .$$

The evaluation of goodness of fit value of (S_{10},S_{20}) with respect to $(\sigma_o^2,\sigma_{\alpha o}^2)$ is simple. We need only to calculate

$$W_o' = \sqrt{(\frac{2S_{10}}{\sigma_o^2})} - \sqrt{[2k(n-1)-1]}$$

$$B_o' = \sqrt{(\frac{2S_{20}}{\sigma_o^2+n\sigma_{\alpha o}^2})} - \sqrt{(2k-3)} \quad .$$

The associated goodness of fit α is given by

$$\alpha = \mathrm{Exp}[-\tfrac{1}{2}(W_o'^2+B_o'^2)] \quad .$$

The contour of equi-goodness of fit level α in (W',B') plane is a circle given by

$$W'^2 + B'^2 = -2 \log \alpha \quad .$$

We can use the Wilson-Hilferty approximation, which is claimed to be better than the Fisher's approximation, but involves more calculations. We need only to re-define W' and B' as under

$$W' = [\{\frac{S_1/k(n-1)}{\sigma^2}\}^{1/3} + \frac{2}{9k(n-1)} - 1][\frac{9k(n-1)}{2}]^{1/2}$$

$$B' = [\{\frac{S_2/(k-1)}{\sigma^2+n\sigma_\alpha^2}\}^{1/3} + \frac{2}{9(k-1)} - 1][\frac{9(k-1)}{2}]^{1/2} \quad .$$

The rest of the procedure is identically the same as explained earlier.

In practice, we rarely have the number of groups, k, large enough to justify a normal approximation of $\chi^2_{k(n-1)}$ as well as χ^2_{k-1}. The utility of the above methods is, therefore, very limited. Quite often, we have k and n such that $k(n-1)$ is large enough to use a normal approximation of $\chi^2_{k(n-1)}$. This case is discussed below.

<u>Asymptotic Case II: k(n-1) large</u> We presume an experimental situation where k is not large enough to justify the normal approximation of χ^2_{k-1}, but k and n are such that $\chi^2_{k(n-1)}$ can be approximated by a normal variate. We can use the Fisher's approximation or the Wilson-Hilferty

approximation. We define (W',B') co-ordinate system as under

$$W' = \sqrt{\left(\frac{2S_1}{\sigma^2}\right)} - \sqrt{[2k(n-1)\ 1]}$$

if the Fisher's approximation is used,

$$= [\{\frac{S_1/k(n-1)}{\sigma^2}\}^{1/3} + \frac{2}{9k(n-1)} -1][\frac{9k(n-1)}{2}]^{1/2}$$

if the Wilson-Hilferty's approximation is used and

$$B' = B = S_2/(\sigma^2+n\sigma_\alpha^2) \quad .$$

Then W' is the standard normal variate and B' is a χ^2_{k-1} variate with $(k-1)$d.f., both being mutually independent. The joint distribution of (W',B') is

$$f(W',B') = [2^{k/2}\pi^{1/2}\Gamma(\frac{k-1}{2})]^{-1}B'^{(k-3)/2}\mathrm{Exp}[-\tfrac{1}{2}(B'+W'^2)]$$

$$B' \geq 0; \ -\infty < W' < \infty \quad .$$

Given $(S_{10},S_{20};\sigma_o^2,\sigma_{\alpha o}^2)$, we can determine uniquely (W_o',B_o') in (W',B') plane. The goodness of fit value associated with (W_o',B_o') is

$$G.F.(W_o',B_o') = \int_\omega [2^{k/2}\pi^{1/2}\ \Gamma(\frac{k-1}{2})\ ]^{-1}B'^{(k-3)/2}\mathrm{Exp}[-\tfrac{1}{2}(B'+W'^2)]dW'dB'$$

where the region of integration ω is given by

$$B'^{\frac{k-3}{2}}\ \mathrm{Exp}[-\tfrac{1}{2}(B'+W'^2)] \leq B_o'^{\frac{k-3}{2}}\ \mathrm{Exp}[-\tfrac{1}{2}(B_o'+W_o'^2)]$$

or

$$(k-3)\log B' - (B'+W'^2) \leq (k-3)\log B_o' - (B_o'+W_o'^2) = C_o' \text{ , say.}$$

The contours of equi-goodness of fit value are given by

$$(k-3)\log B' - (B'+W'^2) = C'_o \quad ,$$

where C'_o is an admissible constant.

The system of contours has the center at $W' = 0$, $B' = (k-3)$ and the maximum value of C'_o is $(k-3)\log(k-3) - (k-3)$ at this point.

The goodness of fit value for any observed (S_{10}, S_{20}) with respect to a given $(\sigma_o^2, \sigma_{\alpha o}^2)$ can be evaluated by the following method.

(i) Calculate $B'_o = \dfrac{S_{10}}{\sigma_o^2 + n\sigma_{\alpha o}^2}$.

(ii) Calculate $W'_o = \sqrt{\left(\dfrac{2S_{10}}{\sigma_o^2}\right)} - \sqrt{[2k(n-1)\ 1]}$

if the Fisher's approximation is used,

$$W'_o = [\{\frac{S_{10}/k(n-1)}{\sigma_o^2}\}^{1/3} + \frac{2}{9k(n-1)} - 1][\frac{9k(n-1)}{2}]^{1/2}$$

if the Wilson-Hilferty's approximation is used.

(iii) Calculate C'_o: $(k-3)\log B'_o - (B'_o + W'^2_o) = C'_o$.

(iv) Evaluate the integral

$$\int_\omega [2^{k/2}\pi^{1/2}\,\Gamma\{(k-1)/2\}]^{-1} B'^{(k-3)/2}\operatorname{Exp}\{-\tfrac{1}{2}(B'+W'^2)\}dW'dB'$$

where the region of integration ω is given by $(k-3)\log B' - (B'+W'^2) \leq C'_o$.

The value of the integral is the goodness of fit value of (S_{10}, S_{20}) with respect to $(\sigma_o^2, \sigma_{\alpha o}^2)$ and can be expressed in percent. The value of C_o' which gives 100 α percent goodness of fit is denoted by C_α'.

It may be noted that n appears in steps (i) and (ii) only. Once we have obtained (W_o', B_o') by using either of the two approximations, we do not need n any longer. The crucial steps (iii) and (iv) depend only on k. A table of goodness of fit at a desired level α will have one entry only viz k.

For practical purposes we need 95 percent and 99 percent goodness of fit values frequently. Given $(S_{10}, S_{20}; \sigma_o^2, \sigma_{\alpha o}^2; n, k)$ we can easily calculate observed C_o and know if the observed C_o falls below (tabulated) $C_{.95}$ or between $C_{.95}$ and $C_{.99}$ or above $C_{.99}$ and say if the fit is poor, good or very good. The concept of goodness of fit does not necessarily involve "accept-reject" rules.

<u>Tabulation of C_α and C_α'</u> For the purpose of illustrations we have tabulated C_α and C_α' values for the following sample sizes, which are the same as used in previous chapters.

(i) k = 4, n = 5 small sample

(ii) k = 10, n = 5 Normal approximation of

(iii) k = 20, n = 5 $\chi^2_{k(n-1)}$ is possible.

The critical C_α and C_α' values of goodness of fit have been calculated for $\alpha = 0.99$ and $\alpha = 0.95$. All the integrals were evaluated, using Monte-Carlo integration on the IBM 360 Model 65 at Iowa State University. Pseudo-random uniform numbers were generated using the IBM scientific sub-routine RANDOU. One exact expression integral required approximately

15,000 Monte Carlo trials for convergence and about 6 seconds of computer time. When $\chi^2_{k(n-1)}$ was approximated by a normal variate, the evaluation of one integral required about 12,000 trials and $5\frac{1}{2}$ seconds of computer time. The critical 95 percent and 99 percent values were obtained by trial and search method. The results are given in Table 8. The contours of C_α for $\alpha = .95$ and 0.99 for k, n pairs (4,5), (10,5) and (20,5) and the contours of C'_α for $\alpha = .95$ and .99 for (k,n) pairs (10,5), (20,5) have been prepared and are given in Figures 32, 33, 34, 35 and 36.

Table 8. 95 percent and 99 percent values of C_α and C'_α

k	n	C_α(Exact method)		C'_α(Approx. method)	
		95 %	99 %	95 %	99%
4	5	21.82	21.92	——	——
10	5	106.74	106.83	6.52	6.60
20	5	292.90	292.97	31.06	31.14

A. Goodness of Fit and Likelihood

Given the sample observations (S_{10}, S_{20}) , we can use goodness of fit as a measure for ordering $(\sigma^2, \sigma^2_\alpha)$. The likelihood is another measure for ordering $(\sigma^2, \sigma^2_\alpha)$. One is, naturally, interested in knowing how these measures behave. In particular, if two pairs $(\sigma^2_1, \sigma^2_{\alpha 1})$ and $(\sigma^2_2, \sigma^2_{\alpha 2})$ have the same likelihood, then will they have the same goodness of fit value and vice-versa? We examine this aspect of the problem.

It may be recalled that we have shown that all (W,B) pairs satisfying the following equation will have the same 100 α percent goodness of fit

$$[k(n-1)-2] \log W + (k-3) \log B - (W+B) = C_\alpha \quad .$$

Writing the above equation in terms of $(S_1,S_2;\sigma^2,\sigma_\alpha^2)$ we have

$$[k(n-1)-2][\log S_1 - \log \sigma^2] + (k-3)[\log S_2 - \log(\sigma^2+n\sigma_\alpha^2)]$$

$$-\left(\frac{S_1}{\sigma^2}+\frac{S_2}{\sigma^2+n\sigma_\alpha^2}\right) = C_\alpha \quad .$$

As long as the quadruplet $(S_1,S_2;\sigma^2,\sigma_\alpha^2)$ satisfies the above equation, we have the same goodness of fit value. This can be achieved by fixing S_1 and S_2 and varying σ^2 and σ_α^2. Hence with (S_{10},S_{20}) as fixed sample observations, the 100 α percent goodness of fit contour in $(\sigma^2,\sigma_\alpha^2)$ plane is given by

$$[k(n-1)-2] \log \sigma^2 + (k-3) \log (\sigma^2+n\sigma_\alpha^2) + \frac{S_{10}}{\sigma^2} + \frac{S_{20}}{\sigma^2+n\sigma_\alpha^2}$$

$$= [k(n-1)-2] \log S_{10} + (k-3) \log S_{20} - C_\alpha$$

$$= G_\alpha \text{ , say.}$$

The striking feature of the system of equi-goodness of fit contours in $(\sigma^2,\sigma_\alpha^2)$ plane is the close resemblance with the system of equi-likelihood contours. If there exists n' such that $k(n-1)-2 = (k-3)(n'-1)$, then the above equation represents an equi-likelihood contour arising from a set of data with $(k-3)$ groups and n' observations per group and the "within" and the "between" sum of squares as S_{10} and S_{20} respectively.

We have drawn equi-likelihood contours of 100 α percent of the maximum likelihood (for various values of α) given by the equation

$$[k(n-1)] \log \sigma^2 + k \log(\sigma^2+n\sigma_\alpha^2) + \frac{S_{10}}{\sigma^2} + \frac{S_{20}}{\sigma^2+n\sigma_\alpha^2} = \text{a constant}$$

$$= L_\alpha \text{ , say.}$$

where L_α is the value of the constant for which the likelihood of $(\sigma_1^2, \sigma_\alpha^2)$ on the contour is 100 α percent of the maximum likelihood.

If we use Thompson's restricted maximum likelihood technique, then 100 α percent of the maximum likelihood contour will be given by

$$[k(n-1)] \log \sigma^2 + (k-1) \log(\sigma^2+n\sigma_\alpha^2) + \frac{S_{10}}{\sigma^2} + \frac{S_{20}}{\sigma^2+n\sigma_\alpha^2} = \text{a constant}$$

$$= L'_\alpha \text{ , say.}$$

In the two equi-likelihood equations, only the multipliers of $\log(\sigma^2+n\sigma_\alpha^2)$ differ by 1 and the rest of terms on the left hand side of equations are the same. With sufficiently large value of k, the difference can be ignored and the two systems of contours coincide.

It is obvious that the system of equi-goodness of fit contours is different from the system of equi-likelihood contours. The systems may have close resemblence in the shape of contours. The pairs of $(\sigma^2, \sigma_\alpha^2)$ with the same goodness of fit value will have different likelihoods and so the pairs with the same likelihood will have different goodness of value. To illustrate this and study the nature of differences, we have used the same data which we have used in previous chapters.

It may be recalled that under each (k,n) case, we have considered 5 sets of data with cumulative F values at 0.25, 0.50, 0.75, 0.95 and 0.99, keeping the same mean square error viz 100 for all k,n and F-values. We have selected three widely apart pairs $(\sigma^2, \sigma_\alpha^2)$ on one selected equi-likelihood contour for each sub-set. The contours have been selected in such a way that for each (k,n) case we have different types of contours and that for each cumulative F-value we have different contours for the three (k,n) cases. Goodness of fit value for the selected $(S_1, S_2; \sigma^2, \sigma_\alpha^2)$ has been calculated by exact method for $(k,n) = (4,5)$, $(10,5)$ and $(20,5)$ and by the Fisher and the Wilson-Hilferty approximation method for $(k,n) = (10,5)$. The observed C_o' values were either identical or close to the second place of decimal for the Fisher and the Helferty approximations for $(k,n) = (20,5)$. We have, therefore, calculated goodness of fit values using the Fisher approximation only. The goodness of fit values to-gether with coded likelihood values are given in Tables 9, 10 and 11. These tables have been prepared with a view to know how approximate goodness of fit values compare with the exact values and to make a comparative study of likelihood and goodness of fit.

A study of Tables 9, 10, and 11 will show for the goodness of fit. values more than 80 percent, normal approximation methods generally compare favourably with the exact method. At some places e.g. for pairs (115,0), (116,60) and (92,71) in Table 10 the differences can not be ignored. It appears that the Wilson-Hilferty approximation has not given any substantial gain over the Fisher's approximation. It may be due to large (40) degrees of freedom for $\chi_{k(n-1)}$. The range of difference in the goodness of fit values of the selected pairs $(\sigma^2, \sigma_\alpha^2)$ with the same likelihood seems to depend on the value of the likelihood as percent of the maximum of likelihood and k. It decreases with an increase in the likelihood value. Perhaps it is due to the fact that points on a contour of large likelihood are not so widely apart as the points on a contour of small likelihood. With large value of k, the range decreases.

Asymptotic case when k is very large   For convenience we reproduce below the equations for contour of equi-likelihood and equi-goodness of fit.

(i) Likelihood used by us

$$[k(n-1)]\log \sigma^2 + k \log(\sigma^2+n\sigma_\alpha^2) + \frac{S_{10}}{\sigma^2} + \frac{S_{20}}{\sigma^2+n\sigma_\alpha^2} = L_\alpha$$

(ii) Thompson's restricted likelihood

$$[k(n-1)]\log \sigma^2 + (k-1)\log(\sigma^2+n\sigma_\alpha^2) + \frac{S_{10}}{\sigma^2} + \frac{S_{20}}{\sigma^2+n\sigma_\alpha^2} = L'_\alpha$$

(iii) Goodness of fit

$$[k(n-1)-2]\log \sigma^2 + (k-3)\log(\sigma^2+n\sigma_\alpha^2) + \frac{S_{10}}{\sigma^2} + \frac{S_{20}}{\sigma^2+n\sigma_\alpha^2} = G_\alpha \quad .$$

If k is so large that the difference between $k(n-1)$ and $[k(n-1)-2]$ and the differences between k, $(k-1)$ and $(k-3)$ can be ignored then we have the same system of contours for (i), (ii) and (iii). For a 100 α percent goodness of fit contour, there exists $\beta(0 \leq B \leq 1)$ such that it is 100 β percent of the maximum likelihood contour. The numbers α and β may be different. The same will hold good a 100 α percent of the maximum likelihood contour. Thus, two pairs $(\sigma_1^2, \sigma_{\alpha 1}^2)$ and $(\sigma_2^2, \sigma_{\alpha 2}^2)$ with the same goodness of fit value will have the same likelihood and vice-versa. As the systems coincide, the ordering will be the same.

Table 9. Goodness of fit for selected equi-likelihood pairs $(\sigma^2, \sigma_\alpha^2)$ by the exact method k = 4, n = 5; S_1 = 1,600

σ^2	σ_α^2	S_2	Contour percent of L_{Max}	Coded[a] Likelihood	Goodness of fit percent
71	2	122	70	109.82	42.79
91	4.8	122	70	109.80	84.44
111	0.8	122	70	109.81	99.76
85	17	248	50	111.91	72.31
66	3	248	50	111.92	22.08
127	6	248	50	111.91	91.11
112	0	454	95	112.71	48.96
90	3	454	95	112.71	36.84
105	7	454	95	112.71	61.67
85	66	972	70	116.37	60.18
80	18	972	70	116.39	17.76
130	42	972	70	116.37	66.62
90	84	1,588	90	117.83	54.05
115	72	1,588	90	117.83	60.88
115	44	1,588	90	117.83	38.66

[a]Coded likelihood = - 2 log(likelihood)- nklogπ.

Table 10. Goodness of fit for selected equi-likelihood pairs $(\sigma^2, \sigma_\alpha^2)$ by exact and approximation methods, $k = 10$, $n = 5$; $S_1 = 4,000$

σ^2	σ_α^2	S_2	Contour $^o/o$ of L_{Max}	Coded[a] Likelihood	Goodness of fit percent: Exact	Fisher Approx.	Wilson-Hilferty Approx.
112	2	582	50	118.16	76.49	71.78	72.37
102	6	582	50	117.98	72.55	71.78	71.78
74	1	582	50	118.15	26.38	28.95	30.03
82	0.4	848	70	279.45	40.18	43.85	43.55
90	6.4	848	70	279.43	77.67	86.63	85.49
115	0.0	848	70	279.40	92.02	85.49	86.09
108	4	1,206	90	282.26	83.72	80.40	80.69
92	3	1,206	90	282.35	51.89	54.84	54.59
98	10	1,206	90	282.34	91.92	94.75	94.55
94	22	1,908	95	286.82	74.28	79.20	78.81
104	14	1,908	95	286.79	67.05	64.79	64.79
96	23.8	1,908	95	286.82	82.69	86.88	86.63
64.1	40	2,592	10	294.40	3.62	7.57	8.79
73.3	100	2,592	10	294.39	16.42	23.75	24.41
159.0	50	2,592	10	294.39	20.87	21.63	19.82

[a]Coded likelihood = - 2 log(likelihood) - nklogπ.

Table 11. Goodness of fit for selected equi-likelihood pairs $(\sigma^2, \sigma_\alpha^2)$ by exact and approximation methods, k = 20, n = 5; S_1 = 8,000

σ^2	σ_α^2	S_2	Contour °/o of L_{Max}	Coded[a] Likelihood	Goodness of fit percent Exact	Fisher's Approx.
100	0.2	1,434	90	554.91	85.88	87.91
97	0.6	1,434	90	554.89	82.69	86.48
90	0.4	1,434	90	554.90	67.11	75.54
103	0.0	1,850	95	559.10	98.42	97.59
95	0.75	1,850	95	559.11	84.61	89.40
100	1.0	1,850	95	559.16	97.50	99.60
88.5	30	2,356	10	568.40	23.91	26.76
72.6	10	2,356	10	568.39	10.05	9.63
113.5	25	2,356	10	568.39	23.23	28.86
105	27	3,286	50	571.81	81.19	80.70
120	5	3,286	50	571.86	39.36	39.80
85	11	3,286	50	571.83	24.88	35.05
110	12	4,086	70	575.52	48.16	50.00
100	34	4,086	70	575.52	91.84	94.56
90	17	4,086	70	575.52	40.40	48.79

[a]Coded likelihood = - 2 log(likelihood) - nklogπ.

Figure 32. Contours of 95 and 99 percent goodness of fit $k = 4$, $n = 5$ 15

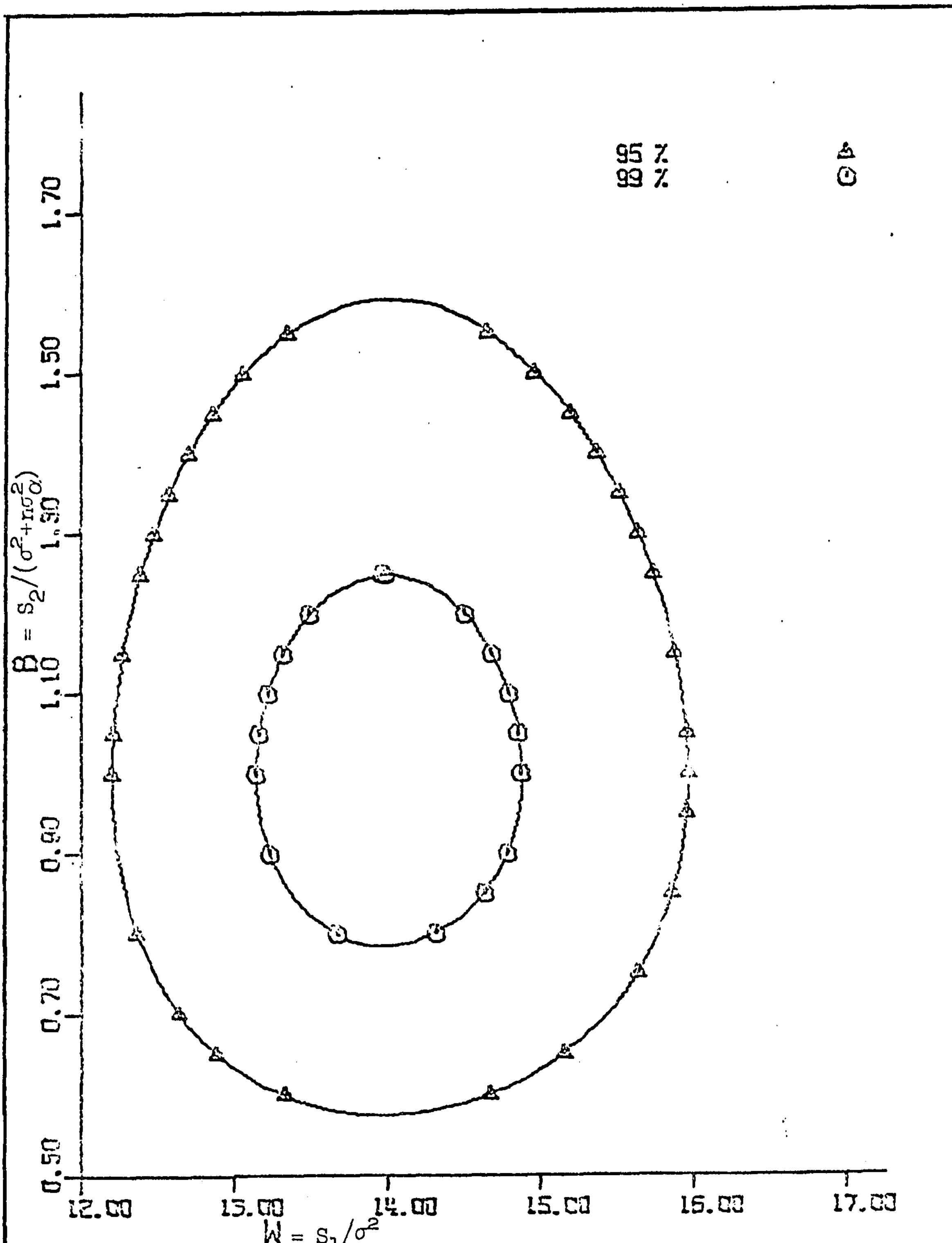

95 %
99 %
B = $S_2/(\sigma^2+n\sigma_\alpha^2)$
0.50
0.70
0.90
1.10
1.30
1.50
1.70
12.00
13.00
14.00
15.00
16.00
17.00
W = S_1/σ^2

Figure 33. Contours of 95 and 99 percent goodness of fit $k = 10$, $n = 5$

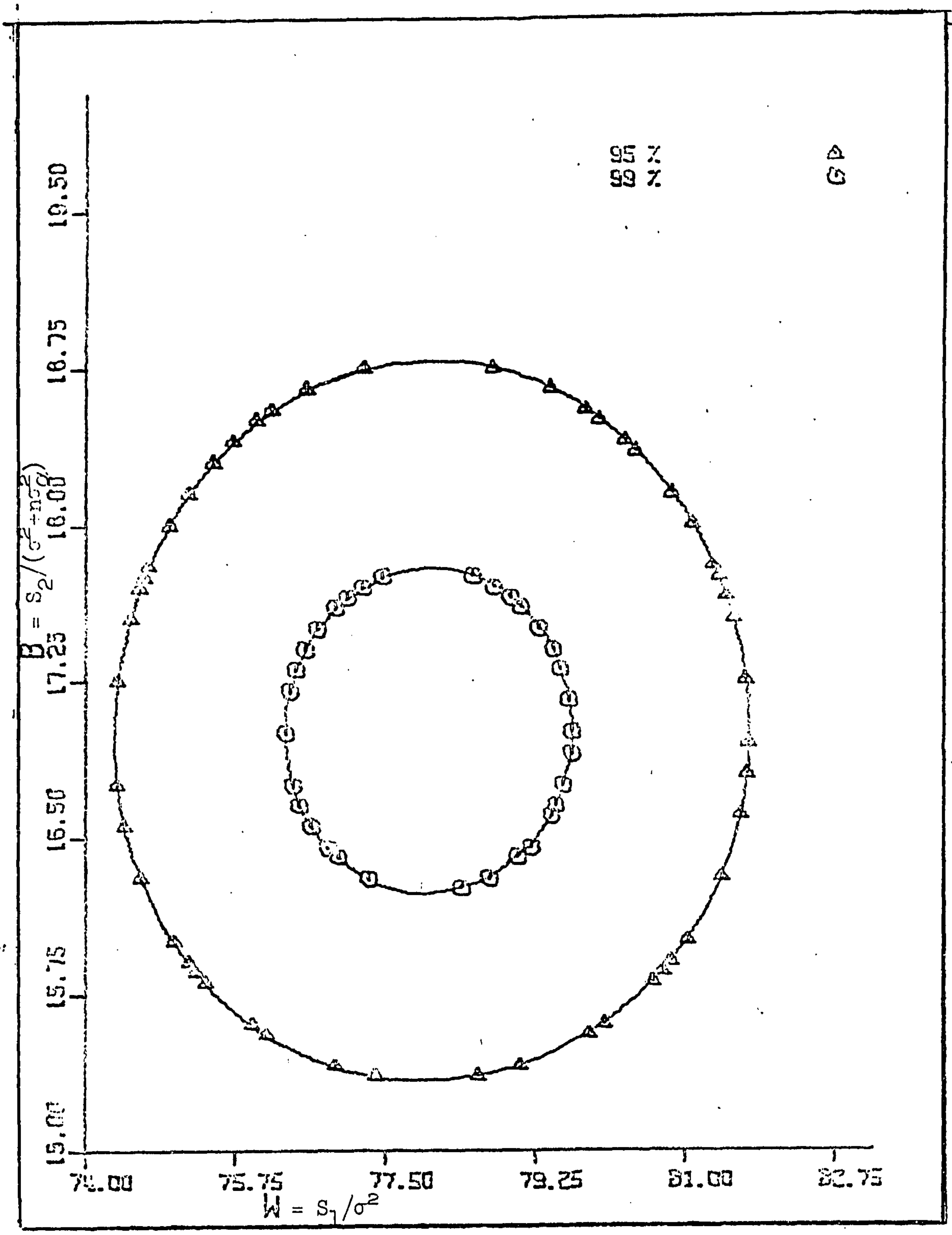

95 %
99 %
$B = S_2/(\sigma^2+n\sigma_g^2)$
19.50
18.75
18.00
17.25
16.50
15.75
15.00
74.00
75.75
77.50
79.25
81.00
82.75
$W = S_1/\sigma^2$

Figure 34. Contours of 95 and 99 percent goodness of fit $k = 20$, $n = 5$

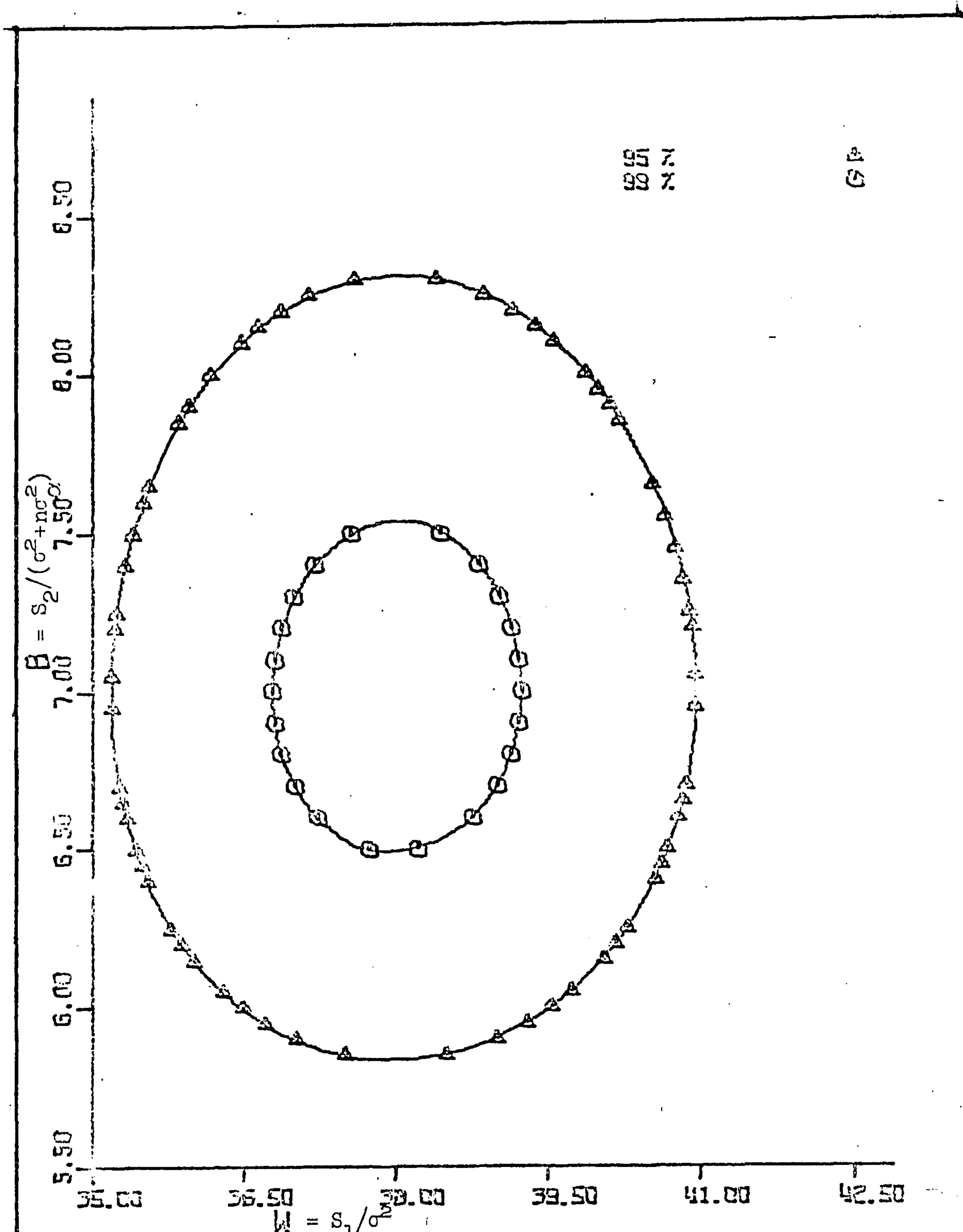

95 %
99 %
8.50
8.00
7.50
7.00
6.50
6.00
5.50
35.00
36.50
38.00
39.50
41.00
42.50
W = S1/σ²

Figure 35. Contours of 95 and 99 percent goodness of fit using normal approximation for error mean square k = 10

$$B = \frac{S_1}{(\sigma^2 + n\sigma_\alpha^2)}$$

$$W' = \sqrt{\left(\frac{2S_1}{\sigma^2}\right)} - \sqrt{[2k(n-1)-1]}$$ for the Fisher's approximation

or

$$= \left[\left\{\frac{S_1/k(n-1)}{\sigma^2}\right\}^{1/3} + \frac{2}{9k(n-1)} - 1\right]\left[\frac{9k(n-1)}{2}\right]^{1/2}$$

for the Wilson-Hilferty's approximation.

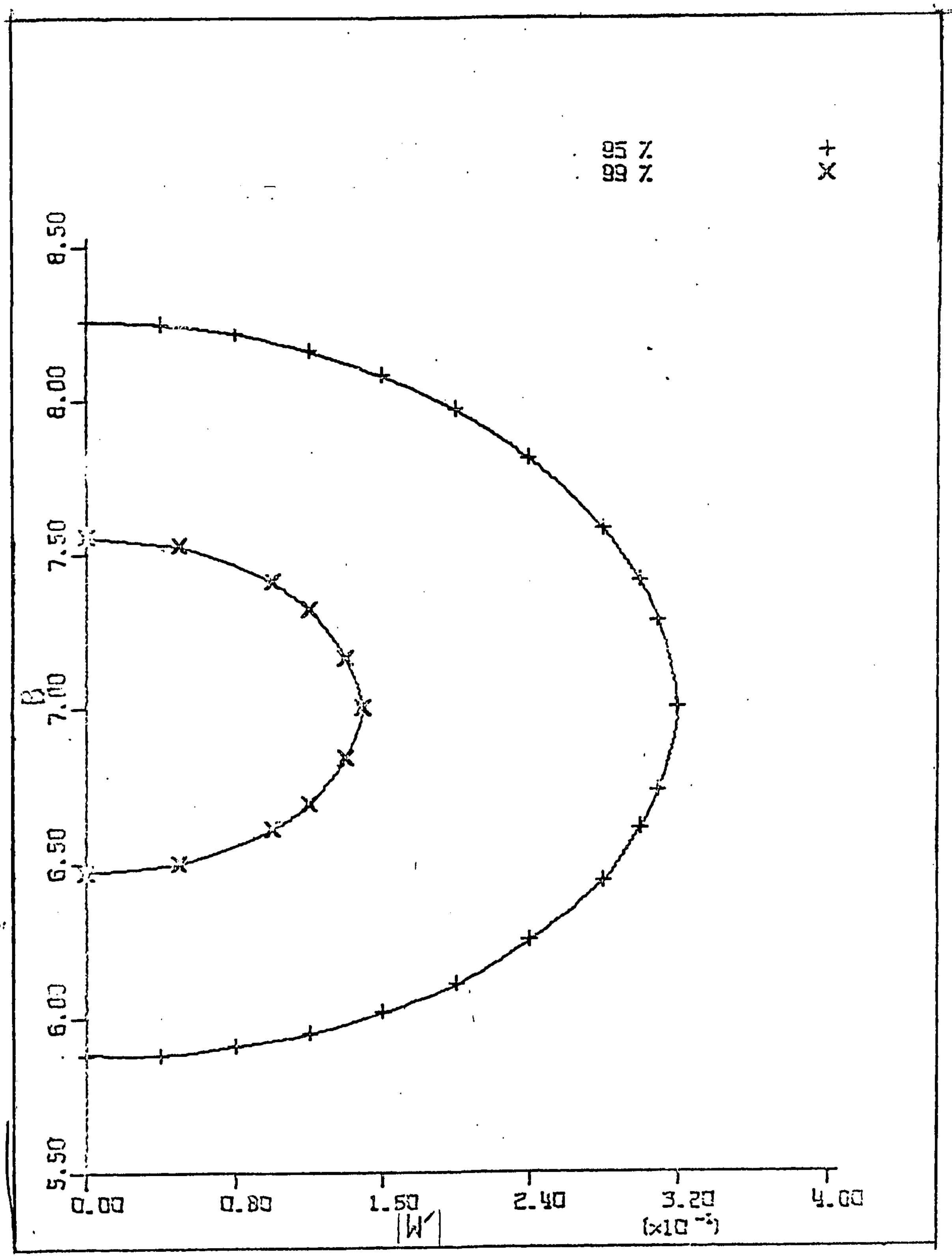

95 %
99 %
8.50
8.00
7.50
7.00
6.50
6.00
5.50
β
0.00
0.80
1.60
2.40
3.20
4.00
|W′|

Figure 36. Contours of 95 and 99 percent goodness of fit using normal approximation for error mean square k = 20

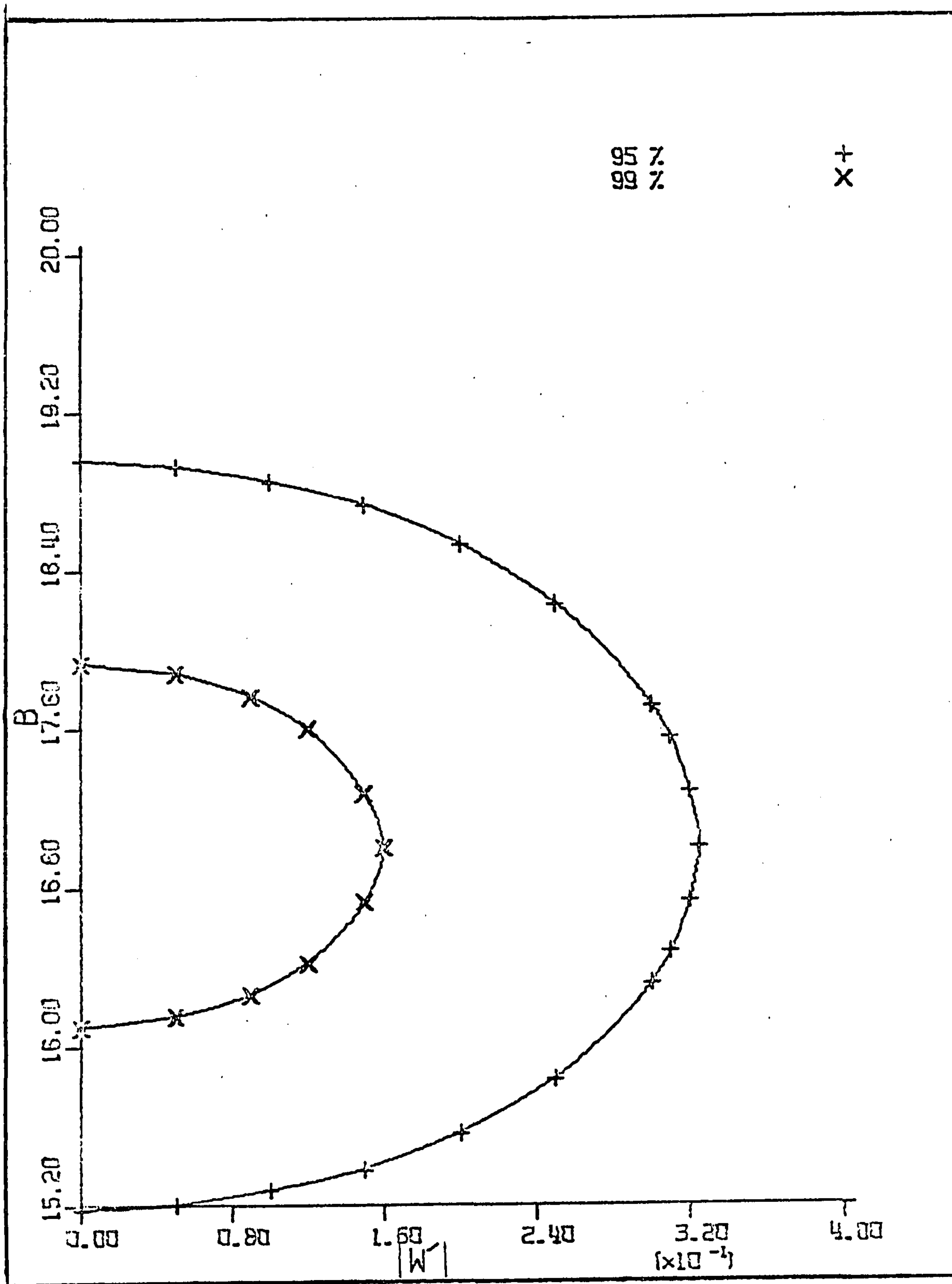
$$B = \frac{S_1}{(\sigma^2+n\sigma_\alpha^2)}$$

$$W' = \sqrt{\left(\frac{2S_1}{\sigma^2}\right)} - \sqrt{[2k(n-1)-1]} \quad \text{for the Fisher's approximation}$$

or

$$= \left[\left\{\frac{S_1/k(n-1)}{\sigma^2}\right\}^{1/3} + \frac{2}{9k(n-1)} - 1\right]\left[\frac{9k(n-1)}{2}\right]^{1/2}$$

for the Wilson-Hilferty's approximation.



95 % +
99 % X
B
15.20
16.00
16.80
17.60
18.40
19.20
20.00
0.00
0.80
1.60
2.40
3.20
4.00
|W'|
(x10 -1)

VI. SUMMARY AND CONCLUSIONS

In the present study we have used the simplest form of components of variance model but the ideas and techniques presented have wide applications. There will be, of course, difficulties of graphic representation and evaluation of complicated integrals. We have approached the problem from three points of view (i) Bayesian analysis (ii) likelihood principle and (iii) goodness of fit.

The particular component of variance situation examined is that of data following the model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$$(i = 1, \dots, k, \; j = 1, \dots, n)$$

$\alpha_i \sim NID(0,\sigma^2_\alpha)$; $\epsilon_{ij} \sim NID(0,\sigma^2)$; all the ϵ_{ij}, α_i mutually independent.

This is called the random one-way classification with k groups and n observations in each group. We define S_1 and S_2 as

$$S_1 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 = \text{'within' sum of squares,}$$

$$S_2 = n \sum_{i=1} (\bar{y}_{i.} - \bar{y}_{..})^2 = \text{'between' sum of squares,}$$

where

$$\bar{y}_{i.} = \sum_{j=1}^{n} y_{ij}/n \; ; \; \bar{y}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}/kn \; .$$

Tiao, Tan and others have used improper priors which contain the sample element n = number of observations in each group. They claim that an improper prior is an expression of lack of knowledge about the parameters and there is no unique way of expression. They have relied on Jeffrey's principle of invariance. Our first objection is against the use of the sample element n . A prior represents the knowledge of an experimenter before the experiment. It should not depend on what he is going to do next. As the prior can be easily absorbed in the likelihood function, subsequent integration for marginal posterior distribtuions is not difficult. We strongly feel that if one has to use an improper prior, it should not depend on a sample element simply because of convenience, when we have modern computers to handle complex problems. Our second objection is against the use of improper priors. We are unable to convince ourselves that the experimenter has absolutely no prior knowledge. He has records of previous work, his own experience and that of coworkers to formulate some beliefs, however vague they may be. These beliefs can be expressed in mathematical form and used as a prior. Of course, this process will usually present difficulties of integration.

We have investigated the properties of posterior distributions of (i) $(\sigma^2+\sigma_\alpha^2)/\sigma^2$ (ii) $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$ using in each case Tiao-Tan-like prior $\frac{d\sigma^2}{\sigma^2}\ \frac{d\tau^2}{\tau^2}$ and Stone-Springer-like prior $\frac{d\sigma}{\tau} \cdot \frac{d\tau}{\tau}$, where $\tau^2 = (\sigma^2+\sigma_\alpha^2)$, instead of $\tau^2 = (\sigma^2+n\sigma_\alpha^2)$ as used by others. We have found that the posterior distributions are either uni-modal or have no mode, depending upon sample constants (k,n,S_1,S_2) . The relation between $(n-1)S_1$ and S_2 plays an important role. Values of mode, mean and A.O.V. estimate have been tabulated for some chosen sets of data. To illustrate the use of an informative prior, we have investigated the properties of posterior distribution of $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$ using the prior distribution of $(\sigma^2+\sigma_\alpha^2)$ as an inverted scalar chi-square $(\ell\chi_\nu^2)$ and the prior distribution of $\sigma_\alpha^2/(\sigma^2+\sigma_\alpha^2)$ as the beta (a,b) distribution. We have found that the posterior distribution can be bi-modal. The relation between kn and (a+b) is important in this connection. The Bayesian analysis has the advantage that the problem of negative estimate of σ^2_α does not exist. An estimate of σ^2_α must be greater than or equal to zero, if the posterior mode is our estimate. If the posterior mean is used for the estimate then it is always positive.

The likelihood function does not factorize into the product of a function of σ^2 alone and a function of σ^2_α alone. The parameters σ^2 , σ^2_α are not separable. We feel that separate investigation of σ^2 and σ^2_α may not be possible and therefore we should examine pairs of $(\sigma^2,\sigma^2_\alpha)$. The likelihood principle can be used to determine relative plausibility of two competing hypotheses about $(\sigma^2,\sigma^2_\alpha)$. We have prepared equi-likelihood contours of 50, 70, 90, 95 and 99 percent of maximum likelihoods for the chosen sets of data. After a study of such contours an experimenter may decide the critical likelihood ratio say 0.50 at which he wishes to compare his results, as he does in a significance test. The plausibility of his hypothesized values of σ^2 , σ^2_α can be, then, determined. As the posterior is proportional to the product of likelihood and prior, the likelihood function plays an important role in a Bayesian analysis. We have studied the effect of non-informative priors (used by us and others) on the likelihood through graphic representation.

We have defined a concept of goodness of fit of data, condensed in (S_{10},S_{20}), with respect to an hypothesized pair $(\sigma^2,\sigma^2_\alpha)$. Given (S_{10},S_{20}), we can plot this against $(\sigma^2,\sigma^2_\alpha)$ in $(\sigma^2,\sigma^2_\alpha)$ plane and have contours of equal goodness of fit value. Working in the coordinate system defined by $W = \frac{S_1}{\sigma^2}$, $B = \frac{S_2}{\sigma^2+n\sigma^2_\alpha}$, we have found that contours of equal goodness of fit value in (W,B) plane are given by

$$[k(n-1)-2]\log W + (k-3)\log B - (W+B) = C_\alpha$$

where C_α is a constant for 100 α percent goodness of fit. We have presented contours of 95 percent and 99 percent goodness of fit and tabulated critical values C_α for (k,n) = (4,5), (10,5) and (20,5). The asymptotic case for k(n-1) large enough to justify a normal approximation of $\chi^2_{k(n-1)}$ is considered. In suitable coordinate system, contours of 95 percent and 99 percent goodness of fit are presented and critical values denoted by C'_α, tabulated for (k,n) = (10,5), (20,5). The asymptotic case, when k is very large, is also considered.

Given (S_{10},S_{20}), goodness of fit values can be used as a measure for ordering $(\sigma^2,\sigma^2_\alpha)$. It may be remembered that the likelihood fraction is is also a means of ordering $(\sigma^2,\sigma^2_\alpha)$. The system of contours of equi-goodness of fit in $(\sigma^2,\sigma^2_\alpha)$ plane for a given (S_{10},S_{20}) closely resembles the corresponding system of equi-likelihood contours as far as the shape of contours is concerned. The system has different centres and therefore two pairs of $(\sigma^2,\sigma^2_\alpha)$, having same goodness of fit value will not have the same likelihood and vice-versa. According to the likelihood principle two pairs of $(\sigma^2,\sigma^2_\alpha)$ having the same likelihood are equally plausible and there is no basis for discriminating between them. Thus, all pairs $(\sigma^2,\sigma^2_\alpha)$ on an equi-likelihood contour are equally plausible. Similarly all pairs on an equi-goodness of fit contour have the same goodness of fit value and there is nothing to discriminate. For very high values of k , the two systems of contours viz equi-likelihood and equi-goodness of fit would coincide. A contour of 100 α percent goodness of fit is also a contour of 100 β percent of maximum likelihood for some $\beta(0 < B \leq 1)$ and vice-versa for large values of k and n .

The difficulty with the likelihood function is that any chosen percentage such as 50 percent used to give a contour of equal likelihood does not have any probability interpretation. The goodness of fit contours on the other hand have a probability interpretation in that, the probability of obtaining a set of data which fits a pair of parameter values as well as or better than the actual set of data is equal to the number associated with the contour.

One can conclude perhaps that a reasonable interpretation of the data should be made on the basis of both the likelihood and the goodness of fit. In small sets of data, parametric values are acceptable only if they give a sizeable proportion of the maximum likelihood and also have a reasonable goodness of fit value.

VII. LITERATURE CITED

1. Anderson, R. L. Negative variance estimates, Query No. 9. Technometrics 7: 75-76. 1965.

2. Barnard G. A., Jenkins, G. M. and Winsten, C. B. Likelihood inference and time series. Royal Statistical Society Journal, Series A, 125: 321-372. 1962.

3. Bartlett, M. S. Approximate confidence intervals I. Biometrika 40: 12-19. 1953.

4. Bartlett, M. S. Approximate confidence intervals II. More than one unknown parameter. Biometrika 40: 306-317. 1953.

5. Birnbaum, Allen. On the foundations of statistical inference. American Statistical Association Journal 57: 269-326. 1962.

6. Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems. Annals of Mathematical Statistics 25: 290-330. 1954.

7. Box, G. E. P. and Tiao, G. C. Bayesian estimation of means for random effect models. American Statistical Association Journal 63: 173-181. 1968.

8. Bross, I. Fiducial interval for variance components. Biometrics 6: 136-144. 1950.

9. Bulmer, M. G. Approximate confidence limits for components of variance. Biometrika 44: 159-167. 1957.

10. Crump, S. L. Estimation of variance components in the analysis of variance. Biometrics Bulletin 2: 7-11. 1946.

11. Crump, S. L. The present status of variance component analysis. Biometrics 7: 1-15. 1951.

12. Daniel, H. E. The estimation of components of variance. Royal Statistical Society Journal, Supplement 6: 168-197. 1939.

13. Fisher, R. A. The fiducial argument in statistical inference. Annals of Eugenics 6: 391-398. 1935.

14. Fisher, R. A. and Yates, F. Statistical tables. Sixth edition. New York, New York, Hafner Publishing Company Incorporated. 1963.

15. Green, J. R. A confidence interval for variance components. Annals of Mathematical Statistics 25: 671-685. 1954.

16. Hacking, Ian. Logic of statistical inference. Cambridge, England. Cambridge University Press. 1965.

17. Healy, M. J. R. Fiducial limits of variance components. Royal Statistical Society Journal, Series B, 25: 128-130. 1963.

18. Henderson, C. R. Estimation of genetic parameters (abstract). Annals of Mathematical Statistics 21: 308. 1950.

19. Hill, B. M. Correlated errors in random-models. American Statistical Association Journal 62: 1387-1400. 1967.

20. Hill, B. M. Inference about variance component in one-way model. American Statistical Association Journal 60: 806-825. 1965.

21. Huitson, A. A method of assigning confidence limits to linear combination of variances. Biometrika 42: 471-479. 1955.

22. Kempthorne, O. The design and analysis of experiments. New York, New York, John Wiley and Sons Incorporated. 1952.

23. Klotz, J. H. and Milton, R. C. Estimates of components of variance with small mean square. Technical Report 36. Madison, Wisconsin, Department of Statistics, University of Wisconsin. 1967.

24. Leone, F. C. and Nelson, L S. Sampling distribution of variance components. I. Empirical studies of balanced nested designs. Technometrics 8: 457-468. 1966.

25. Leone, F. C., Nelson, L. S. and Johnson, N. L. Variance estimates in balanced and unbalanced nested designs (abstract). American Statistical Association Meetings: 56. 1967.

26. Lucas, H. L. A method of estimating components of variance in disproportionate numbers (abstract). American Statistical Association Journal 21: 302. 1950.

27. Neyman, J. and Pearson, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference, Part I. Biometrika 20: 175-240. 1933.

28. Neyman, J. and Pearson, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference, Part II. Biometrika 20: 263-294. 1933.

29. Satterthwaite, F. E. Synthesis of variances. Psychometrika 6: 309-316. 1941.

30. Snedecor, G. W. and Cochran, W. G. Statistical methods. Sixth edition. Ames, Iowa, Iowa State University Press. 1967.

31. Stone, M. and Springer, B. G. F. A paradox involving quasi prior distributions. Biometrika 52: 623-627. 1965.

32. Thompson, W. A., Jr. The problem of negative estimates of variance components. Annals of Mathematical Statistics 33: 273-289. 1962.

33. Thompson, W. A., Jr. and Moore, J. R. Non-negative estimates of variance components. Technometrics 5: 441-449. 1963.

34. Tiao, G.C. and Box, G. E. P. Bayesian analysis of three-component hierarchal design model. Biometrika 54: 109-125. 1967.

35. Tiao, G. C. and Tan, P. Y. Bayesian analysis of random-effect models in the analysis of variance I: posterior distribution of variance-components. Biometrika 52: 37-53. 1965.

36. Tiao, G. C. and Tan, P. Y. Bayesian analysis of random-effect models II: Effect of auto-correlated errors. Biometrika 53: 477-495. 1966.

37. Wald, A. A. On analysis of variance in case of multiple classification with unequal class frequencies. Annals of Mathematical Statistics 12: 346-349. 1941.

38. Wald, A. A. A note on analysis of variance in case of multiple classification with unequal class frequencies. Annals of Mathematical Statistics 11: 96-100. 1940.

39. Welch, B. L. On linear combination of several variances. American Statistical Association Journal 51: 132-148. 1956.

40. Wilk, M. B. and Kempthorne, O. Fixed, mixed and random models. American Statistical Association Journal 50: 1144-1167. 1955.

41. Williams, J. S. A confidence interval for variance components. Biometrika 49: 278-281. 1962.

42. Yates, F. The analysis of multiple classification with unequal numbers in different classes. American Statistical Association Journal 29: 51-56. 1934.

43. Zacks, S. More efficient estimators of variance. Technical report No. 4. Manhattan, Kansas, Department of Statistics, Kansas State University. 1967.

VIII. APPENDIX

A. Artificial Analysis of Variance Data

Sources	d.f.	Cumulative F-values 0.25 Sum of Squares	0.50 Sum of Squares	0.75 Sum of Squares	0.95 Sum of Squares	0.99 Sum of Squares
				(i) $k = 4$, $n = 5$		
Between (S_2)	3	122	248	454	972	1,588
Within (S_1)	16	1,600	1,600	1,600	1,600	1,600
				(ii) $k = 10$, $n = 5$		
Between (S_2)	9	582	848	1,206	1,908	2,592
Within (S_1)	40	4,000	4,000	4,000	4,000	4,000
				(iii) $k = 20$, $n = 5$		
Between (S_2)	19	1,434	1,850	2,356	3,268	4,086
Within (S_1)	80	8,000	8,000	8,000	8,000	8,000

IX. ACKNOWLEDGMENTS

The author wishes to express his deepest gratitude to Professor Oscar Kempthorne for his continued encouragement and guidance throughout the period of study at Iowa State University.

Thanks are due to Messrs. Eric West and Leon Jordan Filho for their valuable assistance in computer programming and to Mrs. Shirley Saveraid for her fine typing of the final manuscript.

The author acknowledges the financial support by the Agency for International Development and the Government of West Pakistan for the completion of his study. A part of the research was supported by Aerospace Research Laboratories, Wright-Patterson Air Force Base, which is gratefully acknowledged.

Finally, the author wishes to thank his wife for her patience and courage during the long absence of the author abroad for his studies. This allowed the author to concentrate fully on his studies.